

---

# **INDRA Documentation**

*Release 1.13.0*

**B. M. Gyori, J. A. Bachman**

**Aug 23, 2019**



|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>License and funding</b>  | <b>3</b>  |
| <b>2</b> | <b>Installation</b>   | <b>5</b>  |
| 2.1      | Installing Python . . . . .   | 5         |
| 2.2      | Installing INDRA . . . . .  | 5         |
| 2.2.1    | Installing via Github . . . . .   | 5         |
| 2.2.2    | Cloning the source code from Github . . . . .                             | 6         |
| 2.2.3    | Installing releases with pip . . . . .                                    | 6         |
| 2.3      | INDRA dependencies . . . . .  | 6         |
| 2.3.1    | PySB and BioNetGen . . . . .  | 6         |
| 2.3.2    | Pyjnius . . . . .   | 6         |
| 2.3.3    | Graphviz . . . . .  | 7         |
| 2.3.4    | Matplotlib . . . . .  | 7         |
| 2.3.5    | Optional additional dependencies . . . . .                                | 7         |
| 2.4      | Configuring INDRA . . . . .   | 8         |
| <b>3</b> | <b>Getting started with INDRA</b>   | <b>9</b>  |
| 3.1      | Importing INDRA and its modules . . . . .                                 | 9         |
| 3.2      | Basic usage examples . . . . .  | 9         |
| 3.2.1    | Reading a sentence with TRIPS . . . . .                                   | 9         |
| 3.2.2    | Reading a PubMed Central article with REACH . . . . .                     | 10        |
| 3.2.3    | Getting the neighborhood of proteins from the BEL Large Corpus . . . . .  | 10        |
| 3.2.4    | Getting paths between two proteins from PathwayCommons (BioPAX) . . . . . | 10        |
| 3.2.5    | Constructing INDRA Statements manually . . . . .                          | 10        |
| 3.2.6    | Assembling a PySB model and exporting to SBML . . . . .                   | 11        |
| <b>4</b> | <b>INDRA modules reference</b>  | <b>13</b> |
| 4.1      | INDRA Statements ( <code>indra.statements</code> ) . . . . .              | 13        |
| 4.1.1    | JSON serialization of INDRA Statements . . . . .                          | 15        |
| 4.2      | Processors for knowledge input ( <code>indra.sources</code> ) . . . . .   | 36        |
| 4.2.1    | Biology-oriented Reading Systems . . . . .                                | 36        |
| 4.2.2    | General Purpose Reading Systems . . . . .                                 | 59        |
| 4.2.3    | Standard Molecular Pathway Databases . . . . .                            | 67        |
| 4.2.4    | Custom Knowledge Bases . . . . .  | 82        |
| 4.3      | Database clients ( <code>indra.databases</code> ) . . . . .               | 90        |
| 4.3.1    | HGNC client ( <code>indra.hgnc_client</code> ) . . . . .                  | 90        |
| 4.3.2    | Uniprot client ( <code>indra.databases.uniprot_client</code> ) . . . . .  | 92        |

|        |   |     |
|--------|---|-----|
| 4.3.3  | ChEBI client ( <code>indra.databases.chebi_client</code> )  | 92  |
| 4.3.4  | BioGRID client ( <code>indra.databases.biogrid_client</code> )  | 93  |
| 4.3.5  | Cell type context client ( <code>indra.databases.context_client</code> )  | 94  |
| 4.3.6  | Network relevance client ( <code>indra.databases.relevance_client</code> )                                      | 94  |
| 4.3.7  | NDEX client ( <code>indra.databases.ndex_client</code> )  | 95  |
| 4.3.8  | cBio portal client ( <code>indra.databases.cbio_client</code> )   | 96  |
| 4.3.9  | ChEMBL client ( <code>indra.databases.chembl_client</code> )  | 99  |
| 4.3.10 | LINCS client ( <code>indra.databases.lincs_client</code> )  | 101 |
| 4.3.11 | MeSH client ( <code>indra.databases.mesh_client</code> )  | 101 |
| 4.3.12 | GO client ( <code>indra.databases.go_client</code> )  | 102 |
| 4.3.13 | PubChem client ( <code>indra.databases.pubchem_client</code> )  | 103 |
| 4.3.14 | miRBase client ( <code>indra.databases.mirbase_client</code> )  | 103 |
| 4.3.15 | Experimental Factor Ontology (EFO) client ( <code>indra.databases.efo_client</code> )                           | 104 |
| 4.3.16 | Human Phenotype Ontology (HP) client ( <code>indra.databases.hp_client</code> )                                 | 104 |
| 4.3.17 | Disease Ontology (DOID) client ( <code>indra.databases.doid_client</code> )                                     | 105 |
| 4.4    | Literature clients ( <code>indra.literature</code> )  | 105 |
| 4.4.1  | Pubmed client ( <code>indra.literature.pubmed_client</code> )   | 106 |
| 4.4.2  | Pubmed Central client ( <code>indra.literature.pmc_client</code> )  | 108 |
| 4.4.3  | CrossRef client ( <code>indra.literature.crossref_client</code> )   | 109 |
| 4.4.4  | Elsevier client ( <code>indra.literature.elsevier_client</code> )   | 109 |
| 4.4.5  | NewsAPI client ( <code>indra.literature.newsapi_client</code> )   | 111 |
| 4.4.6  | Deft Tools ( <code>indra.literature.adeft_tools</code> )  | 112 |
| 4.5    | Preassembly ( <code>indra.preassembler</code> )   | 113 |
| 4.5.1  | Preassembler ( <code>indra.preassembler</code> )  | 113 |
| 4.5.2  | Hierarchy manager ( <code>indra.preassembler.hierarchy_manager</code> )   | 118 |
| 4.5.3  | Entity grounding curation and mapping ( <code>indra.preassembler.grounding_mapper</code> )                      | 121 |
| 4.5.4  | Site curation and mapping ( <code>indra.preassembler.sitemapper</code> )  | 126 |
| 4.5.5  | Ontology mapping ( <code>indra.preassembler.ontology_mapper</code> )  | 129 |
| 4.6    | Belief Engine ( <code>indra.belief</code> )   | 129 |
| 4.7    | Mechanism Linker ( <code>indra.mechlinker</code> )  | 132 |
| 4.8    | Assemblers of model output ( <code>indra.assemblers</code> )  | 135 |
| 4.8.1  | Executable PySB models ( <code>indra.assemblers.pysb.assembler</code> )   | 135 |
| 4.8.2  | Cytoscape networks ( <code>indra.assemblers.cx.assembler</code> )   | 141 |
| 4.8.3  | Natural language ( <code>indra.assemblers.english.assembler</code> )  | 142 |
| 4.8.4  | Node-edge graphs ( <code>indra.assemblers.graph.assembler</code> )  | 144 |
| 4.8.5  | SIF / Boolean networks ( <code>indra.assemblers.sif.assembler</code> )  | 145 |
| 4.8.6  | MITRE “index cards” ( <code>indra.assemblers.index_card.assembler</code> )                                      | 146 |
| 4.8.7  | SBGN output ( <code>indra.assemblers.sbgn.assembler</code> )  | 147 |
| 4.8.8  | Cytoscape JS networks ( <code>indra.assemblers.cyjs.assembler</code> )  | 148 |
| 4.8.9  | Causal analysis graphs ( <code>indra.assemblers.cag.assembler</code> )  | 149 |
| 4.8.10 | Tabular output ( <code>indra.assemblers.tsv.assembler</code> )  | 150 |
| 4.8.11 | HTML browsing and curation ( <code>indra.assemblers.html.assembler</code> )                                     | 151 |
| 4.8.12 | BMI wrapper for PySB-assembled models ( <code>indra.assemblers.pysb.bmi_wrapper</code> )                        | 153 |
| 4.8.13 | PyBEL graphs ( <code>indra.assemblers.pybel.assembler</code> )  | 156 |
| 4.8.14 | Kami models ( <code>indra.assemblers.kami.assembler</code> )  | 157 |
| 4.8.15 | IndraNet Graphs ( <code>indra.assemblers.indranet</code> )  | 158 |
| 4.9    | Explanation ( <code>indra.explanation</code> )  | 161 |
| 4.9.1  | Check whether a model satisfies a property ( <code>indra.explanation.model_checker</code> )                     | 161 |
| 4.10   | Tools ( <code>indra.tools</code> )  | 169 |
| 4.10.1 | Run assembly components in a pipeline ( <code>indra.tools.assemble_corpus</code> )                              | 169 |
| 4.10.2 | Real-time feedback for assembly ( <code>indra.tools.live_curation</code> )                                      | 180 |
| 4.10.3 | Build a network from a gene list ( <code>indra.tools.gene_network</code> )                                      | 182 |
| 4.10.4 | Build an executable model from a fragment of a large network ( <code>indra.tools.executable_subnetwork</code> ) | 183 |

|          |  |            |
|----------|--|------------|
| 4.10.5   | Build a model incrementally over time ( <code>indra.tools.incremental_model</code> ) . . . . . | 184        |
| 4.10.6   | The RAS Machine ( <code>indra.tools.machine</code> ) . . . . .                                 | 185        |
| 4.10.7   | High-throughput Reading Tools ( <code>indra.tools.reading</code> ) . . . . .                   | 187        |
| 4.11     | Util ( <code>indra.util</code> ) . . . . .   | 202        |
| 4.11.1   | Utilities for using AWS ( <code>indra.util.aws</code> ) . . . . .                              | 202        |
| 4.11.2   | A utility to get the INDRA version ( <code>indra.util.get_version</code> ) . . . . .           | 203        |
| 4.11.3   | Define NestedDict ( <code>indra.util.nested_dict</code> ) . . . . .                            | 203        |
| 4.11.4   | Some shorthands for plot formatting ( <code>indra.util.plot_formatting</code> ) . . . . .      | 204        |
| <b>5</b> | <b>Tutorials</b> . . . . .   | <b>205</b> |
| 5.1      | Using natural language to build models . . . . .   | 205        |
| 5.1.1    | Read INDRA Statements from a natural language string . . . . .                                 | 205        |
| 5.1.2    | Assemble the INDRA Statements into a rule-based executable model . . . . .                     | 206        |
| 5.1.3    | Exporting the model into other common formats . . . . .  | 207        |
| 5.2      | The Statement curation interface . . . . .   | 208        |
| 5.2.1    | Curating a Statement . . . . .   | 208        |
| 5.2.2    | Submitting a Curation . . . . .  | 209        |
| 5.2.3    | Curation Guidelines . . . . .  | 209        |
| 5.3      | Large-Scale Machine Reading with Starcluster . . . . .   | 211        |
| 5.3.1    | Install REACH . . . . .  | 212        |
| 5.3.2    | Install Amazon S3 support . . . . .  | 212        |
| 5.3.3    | Install other dependencies . . . . .   | 213        |
| 5.3.4    | Assemble a Corpus of PMIDs . . . . .   | 213        |
| 5.3.5    | Process the papers with REACH . . . . .  | 213        |
| 5.3.6    | Extract INDRA Statements from the REACH output on S3 . . . . .                                 | 214        |
| 5.3.7    | Running the whole pipeline with one script . . . . .   | 214        |
| 5.4      | Large-Scale Machine Reading with Amazon Batch . . . . .  | 215        |
| 5.4.1    | How it Works . . . . .   | 215        |
| 5.5      | Assembling everything known about a particular gene . . . . .                                  | 216        |
| 5.5.1    | Collect mechanisms from PathwayCommons and the BEL Large Corpus . . . . .                      | 217        |
| 5.5.2    | Collect a list of publications that discuss the gene of interest . . . . .                     | 217        |
| 5.5.3    | Get the full text or abstract corresponding to the publications . . . . .                      | 217        |
| 5.5.4    | Read the content of the publications . . . . .   | 217        |
| 5.5.5    | Combine all statements and run pre-assembly . . . . .  | 218        |
| 5.5.6    | Assemble the statements into a network model . . . . .   | 218        |
| <b>6</b> | <b>REST API</b> . . . . .  | <b>219</b> |
| 6.1      | Installation . . . . .   | 219        |
| 6.2      | Launching the REST service . . . . .   | 219        |
| 6.3      | Documentation . . . . .  | 219        |
| <b>7</b> | <b>Indices and tables</b> . . . . .  | <b>221</b> |
|          | <b>Python Module Index</b> . . . . .   | <b>223</b> |
|          | <b>Index</b> . . . . .   | <b>225</b> |



INDRA (the Integrated Network and Dynamical Reasoning Assembler) assembles information about causal mechanisms into a common format that can be used to build several different kinds of predictive and explanatory models. INDRA was originally developed for molecular systems biology and is currently being generalized to other domains.

In molecular biology, sources of mechanistic information include pathway databases, natural language descriptions of mechanisms by human curators, and findings extracted from the literature by text mining.

Mechanistic information from multiple sources is de-duplicated, standardized and assembled into sets of *Statements* with associated evidence. Sets of Statements can then be used to assemble both executable rule-based models (using [PySB](#)) and a variety of different types of network models.





# CHAPTER 1

---

## License and funding

---

INDRA is made available under the 2-clause [BSD license](#). INDRA was developed with funding from ARO grant W911NF-14-1-0397, “Programmatic modelling for reasoning across complex mechanisms” under the DARPA Big Mechanism program, W911NF-14-1-0391, “Active context” under the DARPA Communicating with Computers program, “Global Reading and Assembly for Semantic Probabilistic World Models” in the DARPA World Modelers program, and the DARPA Automated Scientific Discovery Framework project.



### 2.1 Installing Python

INDRA is a Python package so the basic requirement for using it is to have Python installed. Python is shipped with most Linux distributions and with OSX. INDRA works with both Python 2 and 3 (tested with 2.7 and 3.5).

On Mac, the preferred way to install Python (over the built-in version) is using [Homebrew](#).

```
brew install python
```

On Windows, we recommend using [Anaconda](#) which contains compiled distributions of the scientific packages that INDRA depends on (numpy, scipy, pandas, etc).

### 2.2 Installing INDRA

#### 2.2.1 Installing via Github

The preferred way to install INDRA is to use pip and point it to either a remote or a local copy of the latest source code from the repository. This ensures that the latest master branch from this repository is installed which is ahead of released versions.

To install directly from Github, do:

```
pip install git+https://github.com/sorgerlab/indra.git
```

Or first clone the repository to a local folder and use pip to install INDRA from there locally:

```
git clone https://github.com/sorgerlab/indra.git
cd indra
pip install .
```

## 2.2.2 Cloning the source code from Github

You may want to simply clone the source code without installing INDRA as a system-wide package.

```
git clone https://github.com/sorgerlab/indra.git
```

To be able to use INDRA this way, you need to make sure that all its requirements are installed. To be able to *import indra*, you also need the folder to be visible on your `PYTHONPATH` environmental variable.

## 2.2.3 Installing releases with pip

Releases of INDRA are also available via `PyPI`. You can install the latest released version of INDRA as

```
pip install indra
```

## 2.3 INDRA dependencies

INDRA depends on a few standard Python packages (e.g. `rdflib`, `requests`, `objectpath`). These packages are installed automatically by `pip`.

Below we provide a detailed description of some extra dependencies that may require special steps to install.

### 2.3.1 PySB and BioNetGen

INDRA builds on the `PySB` framework to assemble rule-based models of biochemical systems. The `pysb` python package is installed by the standard install procedure. However, to be able to generate mathematical model equations and to export to formats such as SBML, the `BioNetGen` framework also needs to be installed in a way that is visible to `PySB`. Detailed instructions are given in the [PySB documentation](#).

### 2.3.2 Pyjnius

To be able to use INDRA's BioPAX API and optional offline reading via the REACH and Eidos APIs, an additional package called `pyjnius` is needed to allow using Java/Scala classes from Python. This is only strictly required in these input sources and the rest of INDRA will work without `pyjnius`.

1. Install JDK from Oracle: <https://www.oracle.com/technetwork/java/javase/downloads/index.html>. We recommend using Java 8 (INDRA is regularly tested with Java 8), however, Java 11 is also expected to be compatible, with possible extra configuration steps needed that are not described here.

4. Set `JAVA_HOME` to your JDK home directory, for instance

```
export JAVA_HOME=/Library/Java/JavaVirtualMachines/jdk-11.0.2.jdk/Contents/Home
```

3. Then first install `cython` followed by `pyjnius` (tested with version 1.1.4). These need to be broken up into two sequential calls to `pip install`.

```
pip install cython
pip install pyjnius==1.1.4
```

On Mac, you may need to [install Legacy Java for OSX](#). If you have trouble installing it, you can try the following as an alternative. Edit

```
/Library/Java/JavaVirtualMachines/jdk-11.0.2.jdk/Contents/Info.plist
```

(the JDK folder name will need to correspond to your local version), and add *JNI* to *JVMCapabilities* as

```
...
<dict>
  <key>JVMCapabilities</key>
  <array>
    <string>CommandLine</string>
    <string>JNI</string>
  </array>
...
```

### 2.3.3 Graphviz

Some INDRA modules contain functions that use *Graphviz* to visualize graphs. On most systems, doing

```
pip install pygraphviz
```

works. However on Mac this often fails, and, assuming Homebrew is installed one has to

```
brew install graphviz
pip install pygraphviz --install-option="--include-path=/usr/local/include/graphviz/"
↪ --install-option="--library-path=/usr/local/lib/graphviz"
```

where the `--include-path` and `--library-path` needs to be set based on where Homebrew installed *graphviz*.

### 2.3.4 Matplotlib

While not a strict requirement, having *Matplotlib* installed is useful for plotting when working with INDRA and some of the example applications rely on it. It can be installed as

```
pip install matplotlib
```

### 2.3.5 Optional additional dependencies

Some dependencies of INDRA are only needed by certain submodules or are only used in specialized use cases. These are not installed by default but are listed as “extra” requirements, and can be installed separately using *pip*. An extra dependency list (e.g. one called `extra_list`) can be installed as

```
pip install indra[extra_list]
```

You can also install all extra dependencies by doing

```
pip install indra --install-option="complete"
```

or

```
pip install indra[all]
```

In all of the above, you may replace *indra* with `.` (if you’re in a local copy of the *indra* folder or with the Github URL of the INDRA repo, depending on your installation method. See also the corresponding [pip documentation](#) for more information.

The table below provides the name and the description of each “extra” list of dependencies.

| Extra list name | Purpose   |
|-----------------|---|
| biopax          | BioPAX input processing and Pathway Commons queries |
| bel             | BEL input processing and output assembly            |
| trips_offline   | Offline reading with local instance of TRIPS system |
| reach_offline   | Offline reading with local instance of REACH system |
| eidos_offline   | Offline reading with local instance of Eidos system |
| geneways        | Geneways reader input processing                    |
| sofia           | SOFIA reader input processing                       |
| bbn             | BBN reader input processing                         |
| sbml            | SBML model export through the PySB Assembler        |
| machine         | Running a local instance of a “RAS machine”         |
| explanation     | Finding explanatory paths in rule-based models      |
| aws             | Accessing AWS compute and storage resources         |
| graph           | Assembling into a visualizing Graphviz graphs       |
| plot            | Create and display plots                            |

## 2.4 Configuring INDRA

Various aspects of INDRA, including API keys, dependency locations, and Java memory limits, are parameterized by a configuration file that lives in `~/config/indra/config.ini`. The default configuration file is provided in `indra/resources/default_config.ini`, and is copied to `~/config/indra/config.ini` when INDRA starts if no configuration already exists. Every value in the configuration can also be set as an environment variable: for a given configuration key, INDRA will first check for an environment variable with that name and if not present, will use the value in the configuration file. In other words, an environment variable, when set, takes precedence over the value set in the config file.

Configuration values include:

- REACHPATH: The location of the JAR file containing a local instance of the REACH reading system
- EIDOSPATH: The location of the JAR file containing a local instance of the Eidos reading system
- SPARSERPATH: The location of a local instance of the Sparser reading system (path to a folder)
- DRUMPATH: The location of a local installation of the DRUM reading system (path to a folder)
- NDEX\_USERNAME, NDEX\_PASSWORD: Credentials for accessing the NDEX web service
- ELSEVIER\_API\_KEY, ELSEVIER\_INST\_KEY: Elsevier web service API keys
- BIOGRID\_API\_KEY: API key for BioGRID web service (see <http://wiki.thebiogrid.org/doku.php/biogridrest>)
- INDRA\_DEFAULT\_JAVA\_MEM\_LIMIT: Maximum memory limit for Java virtual machines launched by INDRA
- SITEMAPPER\_CACHE\_PATH: Path to an optional cache (a pickle file) for the SiteMapper’s automatically obtained mappings.

---

## Getting started with INDRA

---

### 3.1 Importing INDRA and its modules

INDRA can be imported and used in a Python script or interactively in a Python shell. Note that similar to some other packages (e.g. `scipy`), INDRA doesn't automatically import all its submodules, so `import indra` is not enough to access its submodules. Rather, one has to explicitly import each submodule that is needed. For example to access the BEL API, one has to

```
from indra.sources import bel
```

Similarly, each model output assembler has its own submodule under `indra.assemblers` with the assembler class accessible at the submodule level, so they can be imported as, for instance,

```
from indra.assemblers.pysb import PysbAssembler
```

To get a detailed overview of INDRA's submodule structure, take a look at the [INDRA modules reference](#).

### 3.2 Basic usage examples

Here we show some basic usage examples of the submodules of INDRA. More complex usage examples are shown in the Tutorials section.

#### 3.2.1 Reading a sentence with TRIPS

In this example, we read a sentence via INDRA's TRIPS submodule to produce an INDRA Statement.

```
from indra.sources import trips
sentence = 'MAP2K1 phosphorylates MAPK3 at Thr-202 and Tyr-204'
trips_processor = trips.process_text(sentence)
```

The `trips_processor` object has a `statements` attribute which contains a list of INDRA Statements extracted from the sentence.

### 3.2.2 Reading a PubMed Central article with REACH

In this example, a full paper from [PubMed Central](#) is processed. The paper's PMC ID is [PMC3717945](#).

```
from indra.sources import reach
reach_processor = reach.process_pmc('3717945')
```

The `reach_processor` object has a `statements` attribute which contains a list of INDRA Statements extracted from the paper.

### 3.2.3 Getting the neighborhood of proteins from the BEL Large Corpus

In this example, we search the neighborhood of the KRAS and BRAF proteins in the BEL Large Corpus.

```
from indra.sources import bel
bel_processor = bel.process_pybel_neighborhood(['KRAS', 'BRAF'])
```

The `bel_processor` object has a `statements` attribute which contains a list of INDRA Statements extracted from the queried neighborhood.

### 3.2.4 Getting paths between two proteins from PathwayCommons (BioPAX)

In this example, we search for paths between the BRAF and MAPK3 proteins in the PathwayCommons databases using INDRA's BioPAX API. Note that this example will only work if all dependencies of the `indra.sources.biopax` module are installed.

See the [Installation instructions](#) for more details.

```
from indra.sources import biopax
proteins = ['BRAF', 'MAPK3']
limit = 2
biopax_processor = biopax.process_pc_pathsbetween(proteins, limit)
```

We passed the second argument `limit = 2`, which defines the upper limit on the length of the paths that are searched. By default the limit is 1. The `biopax_processor` object has a `statements` attribute which contains a list of INDRA Statements extracted from the queried paths.

### 3.2.5 Constructing INDRA Statements manually

It is possible to construct INDRA Statements manually or in scripts. The following is a basic example in which we instantiate a Phosphorylation Statement between BRAF and MAP2K1.

```
from indra.statements import Phosphorylation, Agent
braf = Agent('BRAF')
map2k1 = Agent('MAP2K1')
stmt = Phosphorylation(braf, map2k1)
```



### 3.2.6 Assembling a PySB model and exporting to SBML

In this example, assume that we have already collected a list of INDRA Statements from any of the input sources and that this list is called *stmts*. We will instantiate a `PysbAssembler`, which produces a PySB model from INDRA Statements.

```
from indra.assemblers.pysb import PysbAssembler
pa = PysbAssembler()
pa.add_statements(stmts)
model = pa.make_model()
```

Here the *model* variable is a PySB Model object representing a rule-based executable model, which can be further manipulated, simulated, saved and exported to other formats.

For instance, exporting the model to [SBML](#) format can be done as

```
sbml_model = pa.export_model('sbml')
```

which gives an SBML model string in the *sbml\_model* variable, or as

```
pa.export_model('sbml', file_name='model.sbml')
```

which writes the SBML model into the *model.sbml* file. Other formats for export that are supported include BNGL, Kappa and Matlab. For a full list, see the [PySB export module](#).



## 4.1 INDRA Statements (`indra.statements`)

Statements represent mechanistic relationships between biological agents.

Statement classes follow an inheritance hierarchy, with all Statement types inheriting from the parent class *Statement*. At the next level in the hierarchy are the following classes:

### Open Domain

- *Event*
- *Influence*
- *Association*

### Biological Domain

- *Complex*
- *Modification*
- *SelfModification*
- *RegulateActivity*
- *RegulateAmount*
- *ActiveForm*
- *Translocation*
- *Gef*
- *Gap*
- *Conversion*

There are several types of Statements representing post-translational modifications that further inherit from *Modification*:

- *Phosphorylation*
- *Dephosphorylation*
- *Ubiquitination*
- *Deubiquitination*
- *Sumoylation*
- *Desumoylation*
- *Hydroxylation*
- *Dehydroxylation*
- *Acetylation*
- *Deacetylation*
- *Glycosylation*
- *Deglycosylation*
- *Farnesylation*
- *Defarnesylation*
- *Geranylgeranylation*
- *Degeranylgeranylation*
- *Palmitoylation*
- *Depalmitoylation*
- *Myristoylation*
- *Demyristoylation*
- *Ribosylation*
- *Deribosylation*
- *Methylation*
- *Demethylation*

There are additional subtypes of *SelfModification*:

- *Autophosphorylation*
- *Transphosphorylation*

Interactions between proteins are often described simply in terms of their effect on a protein's "activity", e.g., "Active MEK activates ERK", or "DUSP6 inactivates ERK". These types of relationships are indicated by the *RegulateActivity* abstract base class which has subtypes

- *Activation*
- *Inhibition*

while the *RegulateAmount* abstract base class has subtypes

- *IncreaseAmount*
- *DecreaseAmount*

Statements involve one or more *Concepts*, which, depending on the semantics of the Statement, are typically biological *Agents*, such as proteins, represented by the class *Agent*. (However, `:py:class'Influence'` statements involve two or more `:py:class'Event'` objects, each of which takes a `:py:class'Concept'` as an argument.)

Agents can have several types of context specified on them including

- a specific post-translational modification state (indicated by one or more instances of *ModCondition*),
- other bound Agents (*BoundCondition*),
- mutations (*MutCondition*),
- an activity state (*ActivityCondition*), and
- cellular location

The *active* form of an agent (in terms of its post-translational modifications or bound state) is indicated by an instance of the class *ActiveForm*.

Agents also carry grounding information which links them to database entries. These database references are represented as a dictionary in the *db\_refs* attribute of each Agent. The dictionary can have multiple entries. For instance, INDRA's input Processors produce genes and proteins that carry both UniProt and HGNC IDs in *db\_refs*, whenever possible. FamPlex provides a name space for protein families that are typically used in the literature. More information about FamPlex can be found here: <https://github.com/sorgerlab/famplex>

| Type                    | Database          | Example                  |
|-------------------------|-------------------|--------------------------|
| Gene/Protein            | HGNC              | {'HGNC': '11998'}        |
| Gene/Protein            | UniProt           | {'UP': 'P04637'}         |
| Gene/Protein            | Entrez            | {'EGID': '5583'}         |
| Gene/Protein family     | FamPlex           | {'FPLX': 'ERK'}          |
| Gene/Protein family     | InterPro          | {'IP': 'IPR000308'}      |
| Gene/Protein family     | Pfam              | {'PF': 'PF00071'}        |
| Gene/Protein family     | NextProt family   | {'NXPFA': '03114'}       |
| Chemical                | ChEBI             | {'CHEBI': 'CHEBI:63637'} |
| Chemical                | PubChem           | {'PUBCHEM': '42611257'}  |
| Chemical                | LINCS / HMS-LINCS | {'LINCS': '42611257'}    |
| Metabolite              | HMDB              | {'HMDB': 'HMDB00122'}    |
| Process, location, etc. | GO                | {'GO': 'GO:0006915'}     |
| Process, disease, etc.  | MeSH              | {'MESH': 'D008113'}      |
| General terms           | NCIT              | {'NCIT': 'C28597'}       |
| Raw text                | TEXT              | {'TEXT': 'Nf-kappaB'}    |

The evidence for a given Statement, which could include relevant citations, database identifiers, and passages of text from the scientific literature, is contained in one or more *Evidence* objects associated with the Statement.

### 4.1.1 JSON serialization of INDRA Statements

Statements can be serialized into JSON and deserialized from JSON to allow their exchange in a platform-independent way. We also provide a JSON schema (see <http://json-schema.org> to learn about schemas) in [https://raw.githubusercontent.com/sorgerlab/indra/master/indra/resources/statements\\_schema.json](https://raw.githubusercontent.com/sorgerlab/indra/master/indra/resources/statements_schema.json) which can be used to validate INDRA Statements JSONs.

Some validation tools include:

- **jsonschema** a Python package to validate JSON content with respect to a schema

- **ajv-cli** Available at <https://www.npmjs.com/package/ajv-cli> Install with “npm install -g ajv-cli” and then validate with: `ajv -s statements_schema.json -d file_to_validate.json`. This tool provides more sophisticated and better interpretable output than jsonschema.
- **Web based tools** There are a variety of web-based tools for validation with JSON schemas, including <https://www.jsonschemavalidator.net>

**class** `indra.statements.statements.BoundCondition` (*agent*, *is\_bound=True*)  
Bases: `object`

Identify Agents bound (or not bound) to a given Agent in a given context.

#### Parameters

- **agent** (*Agent*) – Instance of Agent.
- **is\_bound** (*bool*) – Specifies whether the given Agent is bound or unbound in the current context. Default is True.

#### Examples

EGFR bound to EGF:

```
>>> egf = Agent('EGF')
>>> egfr = Agent('EGFR', bound_conditions=[BoundCondition(egf)])
```

BRAF *not* bound to a 14-3-3 protein (YWHAB):

```
>>> ywhab = Agent('YWHAB')
>>> braf = Agent('BRAF', bound_conditions=[BoundCondition(ywhab, False)])
```

**class** `indra.statements.statements.MutCondition` (*position*, *residue\_from*, *residue\_to=None*)  
Bases: `object`

Mutation state of an amino acid position of an Agent.

#### Parameters

- **position** (*str*) – Residue position of the mutation in the protein sequence.
- **residue\_from** (*str*) – Wild-type (unmodified) amino acid residue at the given position.
- **residue\_to** (*str*) – Amino acid at the position resulting from the mutation.

#### Examples

Represent EGFR with a L858R mutation:

```
>>> egfr_mutant = Agent('EGFR', mutations=[MutCondition('858', 'L', 'R')])
```

**class** `indra.statements.statements.ModCondition` (*mod\_type*, *residue=None*, *position=None*, *is\_modified=True*)  
Bases: `object`

Post-translational modification state at an amino acid position.

#### Parameters

- **mod\_type** (*str*) – The type of post-translational modification, e.g., ‘phosphorylation’. Valid modification types currently include: ‘phosphorylation’, ‘ubiquitination’, ‘sumoylation’, ‘hydroxylation’, and ‘acetylation’. If an invalid modification type is passed an `InvalidModTypeError` is raised.
- **residue** (*str or None*) – String indicating the modified amino acid, e.g., ‘Y’ or ‘tyrosine’. If `None`, indicates that the residue at the modification site is unknown or unspecified.
- **position** (*str or None*) – String indicating the position of the modified amino acid, e.g., ‘202’. If `None`, indicates that the position is unknown or unspecified.
- **is\_modified** (*bool*) – Specifies whether the modification is present or absent. Setting the flag specifies that the Agent with the `ModCondition` is unmodified at the site.

## Examples

Doubly-phosphorylated MEK (MAP2K1):

```
>>> phospho_mek = Agent('MAP2K1', mods=[
... ModCondition('phosphorylation', 'S', '202'),
... ModCondition('phosphorylation', 'S', '204')])
```

ERK (MAPK1) unphosphorylated at tyrosine 187:

```
>>> unphos_erk = Agent('MAPK1', mods=(
... ModCondition('phosphorylation', 'Y', '187', is_modified=False)))
```

**class** `indra.statements.statements.ActivityCondition` (*activity\_type, is\_active*)

Bases: `object`

An active or inactive state of a protein.

## Examples

Kinase-active MAP2K1:

```
>>> mek_active = Agent('MAP2K1',
... activity=ActivityCondition('kinase', True))
```

Transcriptionally inactive FOXO3:

```
>>> foxo_inactive = Agent('FOXO3',
... activity=ActivityCondition('transcription', False))
```

## Parameters

- **activity\_type** (*str*) – The type of activity, e.g. ‘kinase’. The basic, unspecified molecular activity is represented as ‘activity’. Examples of other activity types are ‘kinase’, ‘phosphatase’, ‘catalytic’, ‘transcription’, etc.
- **is\_active** (*bool*) – Specifies whether the given activity type is present or absent.

**class** `indra.statements.statements.Statement` (*evidence=None, supports=None, supported\_by=None*)

Bases: `object`

The parent class of all statements.

**Parameters**

- **evidence** (None or *Evidence* or list of *Evidence*) – If a list of Evidence objects is passed to the constructor, the value is set to this list. If a bare Evidence object is passed, it is enclosed in a list. If no evidence is passed (the default), the value is set to an empty list.
- **supports** (list of *Statement*) – Statements that this Statement supports.
- **supported\_by** (list of *Statement*) – Statements supported by this statement.

**agent\_list** (*deep\_sorted=False*)

Get the canonicalized agent list.

**get\_hash** (*shallow=True, refresh=False, matches\_fun=None*)

Get a hash for this Statement.

There are two types of hash, “shallow” and “full”. A shallow hash is as unique as the information carried by the statement, i.e. it is a hash of the *matches\_key*. This means that differences in source, evidence, and so on are not included. As such, it is a shorter hash (14 nibbles). The odds of a collision among all the statements we expect to encounter (well under  $10^8$ ) is  $\sim 10^{-9}$  (1 in a billion). Checks for collisions can be done by using the matches keys.

A full hash includes, in addition to the matches key, information from the evidence of the statement. These hashes will be equal if the two Statements came from the same sentences, extracted by the same reader, from the same source. These hashes are correspondingly longer (16 nibbles). The odds of a collision for an expected less than  $10^{10}$  extractions is  $\sim 10^{-9}$  (1 in a billion).

Note that a hash of the Python object will also include the *uuid*, so it will always be unique for every object.

**Parameters**

- **shallow** (*bool*) – Choose between the shallow and full hashes described above. Default is true (e.g. a shallow hash).
- **refresh** (*bool*) – Used to get a new copy of the hash. Default is false, so the hash, if it has been already created, will be read from the attribute. This is primarily used for speed testing.
- **matches\_fun** (*Optional[function]*) – A function which takes a Statement as argument and returns a string matches key which is then hashed. If not provided the Statement’s built-in *matches\_key* method is used.

**Returns hash** – A long integer hash.**Return type** int**make\_generic\_copy** (*deeply=False*)

Make a new matching Statement with no provenance.

All agents and other attributes besides evidence, belief, supports, and supported\_by will be copied over, and a new uuid will be assigned. Thus, the new Statement will satisfy *new\_stmt.matches(old\_stmt)*.

If *deeply* is set to True, all the attributes will be deep-copied, which is comparatively slow. Otherwise, attributes of this statement may be altered by changes to the new matching statement.

**to\_graph** ()

Return Statement as a networkx graph.

**to\_json** (*use\_sbo=False, matches\_fun=None*)

Return serialized Statement as a JSON dict.

**Parameters**



- **use\_sbo** (*Optional[bool]*) – If True, SBO annotations are added to each applicable element of the JSON. Default: False
- **matches\_fun** (*Optional[function]*) – A custom function which, if provided, is used to construct the matches key which is then hashed and put into the return value. Default: None

**Returns** `json_dict` – The JSON-serialized INDRA Statement.

**Return type** `dict`

**class** `indra.statements.statements.Modification` (*enz, sub, residue=None, position=None, evidence=None*)

Bases: `indra.statements.statements.Statement`

Generic statement representing the modification of a protein.

#### Parameters

- **enz** (`indra.statement.Agent`) – The enzyme involved in the modification.
- **sub** (`indra.statement.Agent`) – The substrate of the modification.
- **residue** (*str or None*) – The amino acid residue being modified, or None if it is unknown or unspecified.
- **position** (*str or None*) – The position of the modified amino acid, or None if it is unknown or unspecified.
- **evidence** (None or *Evidence* or list of *Evidence*) – Evidence objects in support of the modification.

**to\_json** (*use\_sbo=False, matches\_fun=None*)

Return serialized Statement as a JSON dict.

#### Parameters

- **use\_sbo** (*Optional[bool]*) – If True, SBO annotations are added to each applicable element of the JSON. Default: False
- **matches\_fun** (*Optional[function]*) – A custom function which, if provided, is used to construct the matches key which is then hashed and put into the return value. Default: None

**Returns** `json_dict` – The JSON-serialized INDRA Statement.

**Return type** `dict`

**class** `indra.statements.statements.AddModification` (*enz, sub, residue=None, position=None, evidence=None*)

Bases: `indra.statements.statements.Modification`

**class** `indra.statements.statements.RemoveModification` (*enz, sub, residue=None, position=None, evidence=None*)

Bases: `indra.statements.statements.Modification`

**class** `indra.statements.statements.SelfModification` (*enz, residue=None, position=None, evidence=None*)

Bases: `indra.statements.statements.Statement`

Generic statement representing the self-modification of a protein.

#### Parameters

- **enz** (`indra.statement.Agent`) – The enzyme involved in the modification, which is also the substrate.

- **residue** (*str or None*) – The amino acid residue being modified, or None if it is unknown or unspecified.
- **position** (*str or None*) – The position of the modified amino acid, or None if it is unknown or unspecified.
- **evidence** (*None or Evidence or list of Evidence*) – Evidence objects in support of the modification.

**to\_json** (*use\_sbo=False, matches\_fun=None*)  
Return serialized Statement as a JSON dict.

#### Parameters

- **use\_sbo** (*Optional[bool]*) – If True, SBO annotations are added to each applicable element of the JSON. Default: False
- **matches\_fun** (*Optional[function]*) – A custom function which, if provided, is used to construct the matches key which is then hashed and put into the return value. Default: None

**Returns** `json_dict` – The JSON-serialized INDRA Statement.

**Return type** `dict`

```
class indra.statements.statements.Phosphorylation(enz, sub, residue=None, position=None, evidence=None)
Bases: indra.statements.statements.AddModification
```

Phosphorylation modification.

#### Examples

MEK (MAP2K1) phosphorylates ERK (MAPK1) at threonine 185:

```
>>> mek = Agent('MAP2K1')
>>> erk = Agent('MAPK1')
>>> phos = Phosphorylation(mek, erk, 'T', '185')
```

```
class indra.statements.statements.Autophosphorylation(enz, residue=None, position=None, evidence=None)
Bases: indra.statements.statements.SelfModification
```

Intramolecular autophosphorylation, i.e., in *cis*.

#### Examples

p38 bound to TAB1 cis-autophosphorylates itself (see PMID:19155529).

```
>>> tab1 = Agent('TAB1')
>>> p38_tab1 = Agent('P38', bound_conditions=[BoundCondition(tab1)])
>>> autophos = Autophosphorylation(p38_tab1)
```

```
class indra.statements.statements.Transphosphorylation(enz, residue=None, position=None, evidence=None)
Bases: indra.statements.statements.SelfModification
```

Autophosphorylation in *trans*.

Transphosphorylation assumes that a kinase is already bound to a substrate (usually of the same molecular species), and phosphorylates it in an intra-molecular fashion. The `enz` property of the statement must have exactly one `bound_conditions` entry, and we assume that `enz` phosphorylates this molecule. The `bound_neg` property is ignored here.

```
class indra.statements.statements.Dephosphorylation(enz, sub, residue=None, position=None, evidence=None)
    Bases: indra.statements.statements.RemoveModification
    Dephosphorylation modification.
```

### Examples

DUSP6 dephosphorylates ERK (MAPK1) at T185:

```
>>> dusp6 = Agent('DUSP6')
>>> erk = Agent('MAPK1')
>>> dephos = Dephosphorylation(dusp6, erk, 'T', '185')
```

```
class indra.statements.statements.Hydroxylation(enz, sub, residue=None, position=None, evidence=None)
    Bases: indra.statements.statements.AddModification
    Hydroxylation modification.
```

```
class indra.statements.statements.Dehydroxylation(enz, sub, residue=None, position=None, evidence=None)
    Bases: indra.statements.statements.RemoveModification
    Dehydroxylation modification.
```

```
class indra.statements.statements.Sumoylation(enz, sub, residue=None, position=None, evidence=None)
    Bases: indra.statements.statements.AddModification
    Sumoylation modification.
```

```
class indra.statements.statements.Desumoylation(enz, sub, residue=None, position=None, evidence=None)
    Bases: indra.statements.statements.RemoveModification
    Desumoylation modification.
```

```
class indra.statements.statements.Acetylation(enz, sub, residue=None, position=None, evidence=None)
    Bases: indra.statements.statements.AddModification
    Acetylation modification.
```

```
class indra.statements.statements.Deacetylation(enz, sub, residue=None, position=None, evidence=None)
    Bases: indra.statements.statements.RemoveModification
    Deacetylation modification.
```

```
class indra.statements.statements.Glycosylation(enz, sub, residue=None, position=None, evidence=None)
    Bases: indra.statements.statements.AddModification
    Glycosylation modification.
```

```
class indra.statements.statements.Deglycosylation(enz, sub, residue=None, position=None, evidence=None)
    Bases: indra.statements.statements.RemoveModification
```

Deglycosylation modification.

```
class indra.statements.statements.Ribosylation(enz, sub, residue=None, position=None,
                                               evidence=None)
```

Bases: *indra.statements.statements.AddModification*

Ribosylation modification.

```
class indra.statements.statements.Deribosylation(enz, sub, residue=None, position=None,
                                                  evidence=None)
```

Bases: *indra.statements.statements.RemoveModification*

Deribosylation modification.

```
class indra.statements.statements.Ubiquitination(enz, sub, residue=None, position=None,
                                                  evidence=None)
```

Bases: *indra.statements.statements.AddModification*

Ubiquitination modification.

```
class indra.statements.statements.Deubiquitination(enz, sub, residue=None, position=None,
                                                    evidence=None)
```

Bases: *indra.statements.statements.RemoveModification*

Deubiquitination modification.

```
class indra.statements.statements.Farnesylation(enz, sub, residue=None, position=None,
                                                  evidence=None)
```

Bases: *indra.statements.statements.AddModification*

Farnesylation modification.

```
class indra.statements.statements.Defarnesylation(enz, sub, residue=None, position=None,
                                                    evidence=None)
```

Bases: *indra.statements.statements.RemoveModification*

Defarnesylation modification.

```
class indra.statements.statements.Geranylgeranylation(enz, sub, residue=None, position=None,
                                                         evidence=None)
```

Bases: *indra.statements.statements.AddModification*

Geranylgeranylation modification.

```
class indra.statements.statements.Degeranylgeranylation(enz, sub, residue=None, position=None,
                                                         evidence=None)
```

Bases: *indra.statements.statements.RemoveModification*

Degeranylgeranylation modification.

```
class indra.statements.statements.Palmitoylation(enz, sub, residue=None, position=None,
                                                    evidence=None)
```

Bases: *indra.statements.statements.AddModification*

Palmitoylation modification.

```
class indra.statements.statements.Depalmitoylation(enz, sub, residue=None, position=None,
                                                      evidence=None)
```

Bases: *indra.statements.statements.RemoveModification*

Depalmitoylation modification.

```
class indra.statements.statements.Myristoylation(enz, sub, residue=None, position=None,
                                                    evidence=None)
```

Bases: *indra.statements.statements.AddModification*

Myristoylation modification.

**class** `indra.statements.statements.Demyristoylation` (*enz, sub, residue=None, position=None, evidence=None*)  
 Bases: `indra.statements.statements.RemoveModification`

Demyristoylation modification.

**class** `indra.statements.statements.Methylation` (*enz, sub, residue=None, position=None, evidence=None*)  
 Bases: `indra.statements.statements.AddModification`

Methylation modification.

**class** `indra.statements.statements.Demethylation` (*enz, sub, residue=None, position=None, evidence=None*)  
 Bases: `indra.statements.statements.RemoveModification`

Demethylation modification.

**class** `indra.statements.statements.RegulateActivity`  
 Bases: `indra.statements.statements.Statement`

Regulation of activity.

This class implements shared functionality of Activation and Inhibition statements and it should not be instantiated directly.

**to\_json** (*use\_sbo=False, matches\_fun=None*)  
 Return serialized Statement as a JSON dict.

#### Parameters

- **use\_sbo** (*Optional[bool]*) – If True, SBO annotations are added to each applicable element of the JSON. Default: False
- **matches\_fun** (*Optional[function]*) – A custom function which, if provided, is used to construct the matches key which is then hashed and put into the return value. Default: None

**Returns** `json_dict` – The JSON-serialized INDRA Statement.

**Return type** dict

**class** `indra.statements.statements.Inhibition` (*subj, obj, obj\_activity='activity', evidence=None*)  
 Bases: `indra.statements.statements.RegulateActivity`

Indicates that a protein inhibits or deactivates another protein.

This statement is intended to be used for physical interactions where the mechanism of inhibition is not explicitly specified, which is often the case for descriptions of mechanisms extracted from the literature.

#### Parameters

- **subj** (*Agent*) – The agent responsible for the change in activity, i.e., the “upstream” node.
- **obj** (*Agent*) – The agent whose activity is influenced by the subject, i.e., the “downstream” node.
- **obj\_activity** (*Optional[str]*) – The activity of the obj Agent that is affected, e.g., its “kinase” activity.
- **evidence** (None or *Evidence* or list of *Evidence*) – Evidence objects in support of the modification.

**class** `indra.statements.statements.Activation` (*subj, obj, obj\_activity='activity', evidence=None*)  
 Bases: `indra.statements.statements.RegulateActivity`

Indicates that a protein activates another protein.

This statement is intended to be used for physical interactions where the mechanism of activation is not explicitly specified, which is often the case for descriptions of mechanisms extracted from the literature.

#### Parameters

- **subj** (*Agent*) – The agent responsible for the change in activity, i.e., the “upstream” node.
- **obj** (*Agent*) – The agent whose activity is influenced by the subject, i.e., the “downstream” node.
- **obj\_activity** (*Optional[str]*) – The activity of the obj Agent that is affected, e.g., its “kinase” activity.
- **evidence** (None or *Evidence* or list of *Evidence*) – Evidence objects in support of the modification.

#### Examples

MEK (MAP2K1) activates the kinase activity of ERK (MAPK1):

```
>>> mek = Agent('MAP2K1')
>>> erk = Agent('MAPK1')
>>> act = Activation(mek, erk, 'kinase')
```

```
class indra.statements.statements.GtpActivation(subj, obj, obj_activity='activity', evidence=None)
```

Bases: *indra.statements.statements.Activation*

```
class indra.statements.statements.ActiveForm(agent, activity, is_active, evidence=None)
```

Bases: *indra.statements.statements.Statement*

Specifies conditions causing an Agent to be active or inactive.

Types of conditions influencing a specific type of biochemical activity can include modifications, bound Agents, and mutations.

#### Parameters

- **agent** (*Agent*) – The Agent in a particular active or inactive state. The sets of ModConditions, BoundConditions, and MutConditions on the given Agent instance indicate the relevant conditions.
- **activity** (*str*) – The type of activity influenced by the given set of conditions, e.g., “kinase”.
- **is\_active** (*bool*) – Whether the conditions are activating (True) or inactivating (False).

```
to_json (use_sbo=False, matches_fun=None)
```

Return serialized Statement as a JSON dict.

#### Parameters

- **use\_sbo** (*Optional[bool]*) – If True, SBO annotations are added to each applicable element of the JSON. Default: False
- **matches\_fun** (*Optional[function]*) – A custom function which, if provided, is used to construct the matches key which is then hashed and put into the return value. Default: None

**Returns** *json\_dict* – The JSON-serialized INDRA Statement.

**Return type** dict

**class** `indra.statements.statements.HasActivity`(*agent*, *activity*, *has\_activity*, *evidence=None*)

Bases: `indra.statements.statements.Statement`

States that an Agent has or doesn't have a given activity type.

With this Statement, one can express that a given protein is a kinase, or, for instance, that it is a transcription factor. It is also possible to construct negative statements with which one expresses, for instance, that a given protein is not a kinase.

#### Parameters

- **agent** (*Agent*) – The Agent that that statement is about. Note that the detailed state of the Agent is not relevant for this type of statement.
- **activity** (*str*) – The type of activity, e.g., “kinase”.
- **has\_activity** (*bool*) – Whether the given Agent has the given activity (True) or not (False).

**class** `indra.statements.statements.Gef`(*gef*, *ras*, *evidence=None*)

Bases: `indra.statements.statements.Statement`

Exchange of GTP for GDP on a small GTPase protein mediated by a GEF.

Represents the generic process by which a guanosine exchange factor (GEF) catalyzes nucleotide exchange on a GTPase protein.

#### Parameters

- **gef** (*Agent*) – The guanosine exchange factor.
- **ras** (*Agent*) – The GTPase protein.

## Examples

SOS1 catalyzes nucleotide exchange on KRAS:

```
>>> sos = Agent('SOS1')
>>> kras = Agent('KRAS')
>>> gef = Gef(sos, kras)
```

**to\_json** (*use\_sbo=False*, *matches\_fun=None*)

Return serialized Statement as a JSON dict.

#### Parameters

- **use\_sbo** (*Optional[bool]*) – If True, SBO annotations are added to each applicable element of the JSON. Default: False
- **matches\_fun** (*Optional[function]*) – A custom function which, if provided, is used to construct the matches key which is then hashed and put into the return value. Default: None

**Returns** `json_dict` – The JSON-serialized INDRA Statement.

**Return type** dict

**class** `indra.statements.statements.Gap`(*gap*, *ras*, *evidence=None*)

Bases: `indra.statements.statements.Statement`

Acceleration of a GTPase protein's GTP hydrolysis rate by a GAP.

Represents the generic process by which a GTPase activating protein (GAP) catalyzes GTP hydrolysis by a particular small GTPase protein.

#### Parameters

- **gap** (*Agent*) – The GTPase activating protein.
- **ras** (*Agent*) – The GTPase protein.

#### Examples

RASA1 catalyzes GTP hydrolysis on KRAS:

```
>>> rasal = Agent('RASA1')
>>> kras = Agent('KRAS')
>>> gap = Gap(rasal, kras)
```

**to\_json** (*use\_sbo=False, matches\_fun=None*)

Return serialized Statement as a JSON dict.

#### Parameters

- **use\_sbo** (*Optional[bool]*) – If True, SBO annotations are added to each applicable element of the JSON. Default: False
- **matches\_fun** (*Optional[function]*) – A custom function which, if provided, is used to construct the matches key which is then hashed and put into the return value. Default: None

**Returns** **json\_dict** – The JSON-serialized INDRA Statement.

**Return type** dict

**class** `indra.statements.statements.Complex` (*members, evidence=None*)

Bases: `indra.statements.statements.Statement`

A set of proteins observed to be in a complex.

**Parameters** **members** (list of *Agent*) – The set of proteins in the complex.

#### Examples

BRAF is observed to be in a complex with RAF1:

```
>>> braf = Agent('BRAF')
>>> raf1 = Agent('RAF1')
>>> cplx = Complex([braf, raf1])
```

**to\_json** (*use\_sbo=False, matches\_fun=None*)

Return serialized Statement as a JSON dict.

#### Parameters

- **use\_sbo** (*Optional[bool]*) – If True, SBO annotations are added to each applicable element of the JSON. Default: False
- **matches\_fun** (*Optional[function]*) – A custom function which, if provided, is used to construct the matches key which is then hashed and put into the return value. Default: None

**Returns** **json\_dict** – The JSON-serialized INDRA Statement.



**Return type** dict

**class** `indra.statements.statements.Translocation` (*agent*, *from\_location=None*,  
*to\_location=None*, *evidence=None*)

Bases: `indra.statements.statements.Statement`

The translocation of a molecular agent from one location to another.

**Parameters**

- **agent** (*Agent*) – The agent which translocates.
- **from\_location** (*Optional[str]*) – The location from which the agent translocates. This must be a valid GO cellular component name (e.g. “cytoplasm”) or ID (e.g. “GO:0005737”).
- **to\_location** (*Optional[str]*) – The location to which the agent translocates. This must be a valid GO cellular component name or ID.

**to\_json** (*use\_sbo=False*, *matches\_fun=None*)

Return serialized Statement as a JSON dict.

**Parameters**

- **use\_sbo** (*Optional[bool]*) – If True, SBO annotations are added to each applicable element of the JSON. Default: False
- **matches\_fun** (*Optional[function]*) – A custom function which, if provided, is used to construct the matches key which is then hashed and put into the return value. Default: None

**Returns json\_dict** – The JSON-serialized INDRA Statement.

**Return type** dict

**class** `indra.statements.statements.RegulateAmount` (*subj*, *obj*, *evidence=None*)

Bases: `indra.statements.statements.Statement`

Superclass handling operations on directed, two-element interactions.

**to\_json** (*use\_sbo=False*, *matches\_fun=None*)

Return serialized Statement as a JSON dict.

**Parameters**

- **use\_sbo** (*Optional[bool]*) – If True, SBO annotations are added to each applicable element of the JSON. Default: False
- **matches\_fun** (*Optional[function]*) – A custom function which, if provided, is used to construct the matches key which is then hashed and put into the return value. Default: None

**Returns json\_dict** – The JSON-serialized INDRA Statement.

**Return type** dict

**class** `indra.statements.statements.DecreaseAmount` (*subj*, *obj*, *evidence=None*)

Bases: `indra.statements.statements.RegulateAmount`

Degradation of a protein, possibly mediated by another protein.

Note that this statement can also be used to represent inhibitors of synthesis (e.g., cycloheximide).

**Parameters**

- **subj** (`indra.statement.Agent`) – The protein mediating the degradation.

- **obj** (`indra.statement.Agent`) – The protein that is degraded.
- **evidence** (None or *Evidence* or list of *Evidence*) – Evidence objects in support of the degradation statement.

**class** `indra.statements.statements.IncreaseAmount` (*subj, obj, evidence=None*)

Bases: `indra.statements.statements.RegulateAmount`

Synthesis of a protein, possibly mediated by another protein.

#### Parameters

- **subj** (`indra.statement.Agent`) – The protein mediating the synthesis.
- **obj** (`indra.statement.Agent`) – The protein that is synthesized.
- **evidence** (None or *Evidence* or list of *Evidence*) – Evidence objects in support of the synthesis statement.

**class** `indra.statements.statements.Influence` (*subj, obj, evidence=None*)

Bases: `indra.statements.statements.Statement`

An influence on the quantity of a concept of interest.

#### Parameters

- **subj** (`indra.statement.Event`) – The event which acts as the influencer.
- **obj** (`indra.statement.Event`) – The event which acts as the influencee.
- **evidence** (None or *Evidence* or list of *Evidence*) – Evidence objects in support of the statement.

**agent\_list** (*deep\_sorted=False*)

Get the canonicalized agent list.

**to\_json** (*use\_sbo=False, matches\_fun=None*)

Return serialized Statement as a JSON dict.

#### Parameters

- **use\_sbo** (*Optional[bool]*) – If True, SBO annotations are added to each applicable element of the JSON. Default: False
- **matches\_fun** (*Optional[function]*) – A custom function which, if provided, is used to construct the matches key which is then hashed and put into the return value. Default: None

**Returns** `json_dict` – The JSON-serialized INDRA Statement.

**Return type** `dict`

**class** `indra.statements.statements.Conversion` (*subj, obj\_from=None, obj\_to=None, evidence=None*)

Bases: `indra.statements.statements.Statement`

Conversion of molecular species mediated by a controller protein.

#### Parameters

- **subj** (`indra.statement.Agent`) – The protein mediating the conversion.
- **obj\_from** (list of `indra.statement.Agent`) – The list of molecular species being consumed by the conversion.
- **obj\_to** (list of `indra.statement.Agent`) – The list of molecular species being created by the conversion.

- **evidence** (None or *Evidence* or list of *Evidence*) – Evidence objects in support of the synthesis statement.

**to\_json** (*use\_sbo=False, matches\_fun=None*)

Return serialized Statement as a JSON dict.

#### Parameters

- **use\_sbo** (*Optional[bool]*) – If True, SBO annotations are added to each applicable element of the JSON. Default: False
- **matches\_fun** (*Optional[function]*) – A custom function which, if provided, is used to construct the matches key which is then hashed and put into the return value. Default: None

**Returns json\_dict** – The JSON-serialized INDRA Statement.

**Return type** dict

**class** `indra.statements.statements.Unresolved` (*uuid\_str=None, shallow\_hash=None, full\_hash=None*)

Bases: `indra.statements.statements.Statement`

A special statement type used in support when a uuid can't be resolved.

When using the `stmts_from_json` method, it is sometimes not possible to resolve the uuid found in *support* and *supported\_by* in the json representation of an indra statement. When this happens, this class is used as a place-holder, carrying only the uuid of the statement.

**class** `indra.statements.statements.Association` (*members, evidence=None*)

Bases: `indra.statements.statements.Complex`

A set of events associated with each other without causal relationship.

#### Parameters

- **members** (*list of :py:class:Event*) – A list of events associated with each other.
- **evidence** (None or *Evidence* or list of *Evidence*) – Evidence objects in support of the modification.

**agent\_list** (*deep\_sorted=False*)

Get the canonicalized agent list.

**to\_json** (*use\_sbo=False, matches\_fun=None*)

Return serialized Statement as a JSON dict.

#### Parameters

- **use\_sbo** (*Optional[bool]*) – If True, SBO annotations are added to each applicable element of the JSON. Default: False
- **matches\_fun** (*Optional[function]*) – A custom function which, if provided, is used to construct the matches key which is then hashed and put into the return value. Default: None

**Returns json\_dict** – The JSON-serialized INDRA Statement.

**Return type** dict

**class** `indra.statements.statements.Event` (*concept, delta=None, context=None, evidence=None, supports=None, supported\_by=None*)

Bases: `indra.statements.statements.Statement`

An event representing the change of a Concept.

**concept**

The concept over which the event is defined.

**Type** `indra.statements.concept.Concept`

**delta**

Represents a change in the concept, with a polarity and an adjectives entry.

**Type** `indra.statements.delta.Delta`

**context**

The context associated with the event.

**Type** `indra.statements.context.Context`

**to\_json** (*with\_evidence=True, use\_sbo=False, matches\_fun=None*)

Return serialized Statement as a JSON dict.

**Parameters**

- **use\_sbo** (*Optional[bool]*) – If True, SBO annotations are added to each applicable element of the JSON. Default: False
- **matches\_fun** (*Optional[function]*) – A custom function which, if provided, is used to construct the matches key which is then hashed and put into the return value. Default: None

**Returns** `json_dict` – The JSON-serialized INDRA Statement.

**Return type** `dict`

**class** `indra.statements.statements.Migration` (*concept, delta=None, context=None, evidence=None, supports=None, supported\_by=None*)

Bases: `indra.statements.statements.Event`

A special class of Event representing Migration.

**exception** `indra.statements.statements.InputError`

Bases: `Exception`

**exception** `indra.statements.statements.UnresolvedUuidError`

Bases: `Exception`

**exception** `indra.statements.statements.InvalidLocationError` (*name*)

Bases: `ValueError`

Invalid cellular component name.

**exception** `indra.statements.statements.InvalidResidueError` (*name*)

Bases: `ValueError`

Invalid residue (amino acid) name.

**exception** `indra.statements.statements.NotAStatementName`

Bases: `Exception`

**class** `indra.statements.statements.Concept` (*name, db\_refs=None*)

Bases: `object`

A concept/entity of interest that is the argument of a Statement

**Parameters**

- **name** (*str*) – The name of the concept, possibly a canonicalized name.
- **db\_refs** (*dict*) – Dictionary of database identifiers associated with this concept.

```
class indra.statements.statements.Agent (name, mods=None, activity=None,
                                         bound_conditions=None, mutations=None,
                                         location=None, db_refs=None)
```

Bases: `indra.statements.concept.Concept`

A molecular entity, e.g., a protein.

#### Parameters

- **name** (*str*) – The name of the agent, preferably a canonicalized name such as an HGNC gene name.
- **mods** (list of *ModCondition*) – Modification state of the agent.
- **bound\_conditions** (list of *BoundCondition*) – Other agents bound to the agent in this context.
- **mutations** (list of *MutCondition*) – Amino acid mutations of the agent.
- **activity** (*ActivityCondition*) – Activity of the agent.
- **location** (*str*) – Cellular location of the agent. Must be a valid name (e.g. “nucleus”) or identifier (e.g. “GO:0005634”) for a GO cellular compartment.
- **db\_refs** (*dict*) – Dictionary of database identifiers associated with this agent.

```
entity_matches_key ()
```

Return a key to identify the identity of the Agent not its state.

The key is based on the preferred grounding for the Agent, or if not available, the name of the Agent is used.

**Returns** The key used to identify the Agent.

**Return type** `str`

```
get_grounding ()
```

Return a tuple of a preferred grounding namespace and ID.

**Returns** A tuple whose first element is a grounding namespace (HGNC, CHEBI, etc.) and the second element is an identifier in the namespace. If no preferred grounding is available, a tuple of Nones is returned.

**Return type** `tuple`

```
matches_key ()
```

Return a key to identify the identity and state of the Agent.

```
state_matches_key ()
```

Return a key to identify the state of the Agent.

```
class indra.statements.statements.Evidence (source_api=None, source_id=None,
                                             pmid=None, text=None, annotations=None,
                                             epistemics=None, context=None,
                                             text_refs=None)
```

Bases: `object`

Container for evidence supporting a given statement.

#### Parameters

- **source\_api** (*str or None*) – String identifying the INDRA API used to capture the statement, e.g., ‘trips’, ‘biopax’, ‘bel’.
- **source\_id** (*str or None*) – For statements drawn from databases, ID of the database entity corresponding to the statement.

- **pmid** (*str* or *None*) – String indicating the Pubmed ID of the source of the statement.
- **text** (*str*) – Natural language text supporting the statement.
- **annotations** (*dict*) – Dictionary containing additional information on the context of the statement, e.g., species, cell line, tissue type, etc. The entries may vary depending on the source of the information.
- **epistemics** (*dict*) – A dictionary describing various forms of epistemic certainty associated with the statement.
- **text\_refs** (*dict*) – A dictionary of various reference ids to the source text, e.g. DOI, PMID, URL, etc.

There are some attributes which are not set by the parameters above:

**source\_hash** [int] A hash calculated from the evidence text, source api, and pmid and/or source\_id if available. This is generated automatically when the object is instantiated.

**stmt\_tag** [int] This is a hash calculated by a Statement to which this evidence refers, and is set by said Statement. It is useful for tracing ownership of an Evidence object.

**get\_source\_hash** (*refresh=False*)

Get a hash based off of the source of this statement.

The resulting value is stored in the source\_hash attribute of the class and is preserved in the json dictionary.

**to\_json** ()

Convert the evidence object into a JSON dict.

**class** `indra.statements.statements.QualitativeDelta` (*polarity=None, adjectives=None*)

Bases: `indra.statements.delta.Delta`

Qualitative delta defining an Event.

#### Parameters

- **polarity** (*1, -1* or *None*) – Polarity of an Event.
- **adjectives** (*list[str]*) – Adjectives describing an Event.

**class** `indra.statements.statements.QuantitativeState` (*entity=None, value=None, unit=None, modifier=None, text=None*)

Bases: `indra.statements.delta.Delta`

An object representing numerical value of something.

#### Parameters

- **entity** (*str*) – An entity to capture the quantity of.
- **value** (*float* or *int*) – Quantity of a unit (or range?)
- **unit** (*str*) – Measurement unit of value (e.g. absolute, daily, percentage, etc.)
- **modifier** (*str*) – Modifier to value (e.g. more than, at least, approximately, etc.)
- **text** (*str*) – Natural language text describing quantitative state.

**static convert\_unit** (*source\_unit, target\_unit, source\_value, source\_period=None, target\_period=None*)

Convert value per unit from source to target unit. If a unit is absolute, total timedelta period has to be provided. If a unit is a month or a year, it is recommended to pass timedelta period object directly, if not provided, the approximation will be used.

**static from\_seconds** (*value\_per\_second, period*)

Get total value per given period given timedelta period object and value per second.

**static value\_per\_second** (*value, period*)

Get value per second given total value per period and a timedelta period object.

**class** `indra.statements.statements.BioContext` (*location=None, cell\_line=None, cell\_type=None, organ=None, disease=None, species=None*)

Bases: `indra.statements.context.Context`

An object representing the context of a Statement in biology.

#### Parameters

- **location** (*Optional[RefContext]*) – Cellular location, typically a sub-cellular compartment.
- **cell\_line** (*Optional[RefContext]*) – Cell line context, e.g., a specific cell line, like BT20.
- **cell\_type** (*Optional[RefContext]*) – Cell type context, broader than a cell line, like macrophage.
- **organ** (*Optional[RefContext]*) – Organ context.
- **disease** (*Optional[RefContext]*) – Disease context.
- **species** (*Optional[RefContext]*) – Species context.

**class** `indra.statements.statements.WorldContext` (*time=None, geo\_location=None*)

Bases: `indra.statements.context.Context`

An object representing the context of a Statement in time and space.

#### Parameters

- **time** (*Optional[TimeContext]*) – A TimeContext object representing the temporal context of the Statement.
- **geo\_location** (*Optional[RefContext]*) – The geographical location context represented as a RefContext

**class** `indra.statements.statements.TimeContext` (*text=None, start=None, end=None, duration=None*)

Bases: `object`

An object representing the time context of a Statement

#### Parameters

- **text** (*Optional[str]*) – A string representation of the time constraint, typically as seen in text.
- **start** (*Optional[datetime]*) – A *datetime* object representing the start time
- **end** (*Optional[datetime]*) – A *datetime* object representing the end time
- **duration** (*int*) – The duration of the time constraint in seconds

**class** `indra.statements.statements.RefContext` (*name=None, db\_refs=None*)

Bases: `object`

An object representing a context with a name and references.

#### Parameters

- **name** (*Optional[str]*) – The name of the given context. In some cases a text name will not be available so this is an optional parameter with the default being None.
- **db\_refs** (*Optional[dict]*) – A dictionary where each key is a namespace and each value is an identifier in that namespace, similar to the db\_refs associated with Concepts/Agents.

**class** `indra.statements.statements.Context`

Bases: `object`

An abstract class for Contexts.

**class** `indra.statements.statements.MovementContext` (*locations=None, time=None*)

Bases: `indra.statements.context.Context`

An object representing the context of a movement between start and end points in time.

#### Parameters

- **locations** (*Optional[list[dict]]*) – A list of dictionaries each containing a Ref-Context object representing geographical location context and its role (e.g. ‘origin’, ‘destination’, etc.)
- **time** (*Optional[TimeContext]*) – A TimeContext object representing the temporal context of the Statement.

`indra.statements.statements.stmts_from_json` (*json\_in, on\_missing\_support='handle'*)

Get a list of Statements from Statement jsons.

In the case of pre-assembled Statements which have *supports* and *supported\_by* lists, the uuids will be replaced with references to Statement objects from the json, where possible. The method of handling missing support is controlled by the *on\_missing\_support* key-word argument.

#### Parameters

- **json\_in** (*iterable[dict]*) – A json list containing json dict representations of INDRA Statements, as produced by the *to\_json* methods of subclasses of Statement, or equivalently by *stmts\_to\_json*.
- **on\_missing\_support** (*Optional[str]*) – Handles the behavior when a uuid reference in *supports* or *supported\_by* attribute cannot be resolved. This happens because uuids can only be linked to Statements contained in the *json\_in* list, and some may be missing if only some of all the Statements from pre- assembly are contained in the list.

Options:

- ‘handle’: (default) convert unresolved uuids into *Unresolved* Statement objects.
- ‘ignore’: Simply omit any uuids that cannot be linked to any Statements in the list.
- ‘error’: Raise an error upon hitting an un-linkable uuid.

**Returns** `stmts` – A list of INDRA Statements.

**Return type** `list[Statement]`

`indra.statements.statements.get_unresolved_support_uuids` (*stmts*)

Get uuids unresolved in support from stmts from *stmts\_from\_json*.

`indra.statements.statements.stmts_to_json` (*stmts\_in, use\_sbo=False, matches\_fun=None*)

Return the JSON-serialized form of one or more INDRA Statements.

#### Parameters



- **stmts\_in** (*Statement or list[Statement]*) – A Statement or list of Statement objects to serialize into JSON.
- **use\_sbo** (*Optional[bool]*) – If True, SBO annotations are added to each applicable element of the JSON. Default: False
- **matches\_fun** (*Optional[function]*) – A custom function which, if provided, is used to construct the matches key which is then hashed and put into the return value. Default: None

**Returns** `json_dict` – JSON-serialized INDRA Statements.

**Return type** `dict`

`indra.statements.statements.stmts_from_json_file(fname)`

Return a list of statements loaded from a JSON file.

**Parameters** `fname` (*str*) – Path to the JSON file to load statements from.

**Returns** The list of INDRA Statements loaded from the JSON file.

**Return type** `list[indra.statements.Statement]`

`indra.statements.statements.stmts_to_json_file(stmts, fname, **kwargs)`

Serialize a list of INDRA Statements into a JSON file.

**Parameters**

- **stmts** (*list[indra.statement.Statements]*) – The list of INDRA Statements to serialize into the JSON file.
- **fname** (*str*) – Path to the JSON file to serialize Statements into.

`indra.statements.statements.get_valid_residue(residue)`

Check if the given string represents a valid amino acid residue.

`indra.statements.statements.get_valid_location(location)`

Check if the given location represents a valid cellular component.

`indra.statements.statements.get_valid_location(location)`

Check if the given location represents a valid cellular component.

`indra.statements.statements.draw_stmt_graph(stmts)`

Render the attributes of a list of Statements as directed graphs.

The layout works well for a single Statement or a few Statements at a time. This function displays the plot of the graph using `plt.show()`.

**Parameters** `stmts` (*list[indra.statements.Statement]*) – A list of one or more INDRA Statements whose attribute graph should be drawn.

`indra.statements.statements.get_all_descendants(parent)`

Get all the descendants of a parent class, recursively.

`indra.statements.statements.make_statement_camel(stmt_name)`

Makes a statement name match the case of the corresponding statement.

`indra.statements.statements.get_statement_by_name(stmt_name)`

Get a statement class given the name of the statement class.

`indra.statements.statements.make_hash(s, n_bytes)`

Make the hash from a matches key.

`indra.statements.statements.stmt_type(obj, mk=True)`

Return standardized, backwards compatible object type String.

This is a temporary solution to make sure type comparisons and matches keys of Statements and related classes are backwards compatible.

## 4.2 Processors for knowledge input (`indra.sources`)

INDRA interfaces with and draws knowledge from many sources including reading systems (some that extract biological mechanisms, and some that extract general causal interactions from text) and also from structured databases, which are typically human-curated or derived from experimental data.

### 4.2.1 Biology-oriented Reading Systems

#### REACH (`indra.sources.reach`)

##### REACH API (`indra.sources.reach.api`)

Methods for obtaining a reach processor containing indra statements.

Many file formats are supported. Many will run reach.

`indra.sources.reach.api.process_json_file` (*file\_name*, *citation=None*)

Return a ReachProcessor by processing the given REACH json file.

The output from the REACH parser is in this json format. This function is useful if the output is saved as a file and needs to be processed. For more information on the format, see: <https://github.com/clulab/reach>

##### Parameters

- **file\_name** (*str*) – The name of the json file to be processed.
- **citation** (*Optional[str]*) – A PubMed ID passed to be used in the evidence for the extracted INDRA Statements. Default: None

**Returns** **rp** – A ReachProcessor containing the extracted INDRA Statements in `rp.statements`.

**Return type** *ReachProcessor*

`indra.sources.reach.api.process_json_str` (*json\_str*, *citation=None*)

Return a ReachProcessor by processing the given REACH json string.

The output from the REACH parser is in this json format. For more information on the format, see: <https://github.com/clulab/reach>

##### Parameters

- **json\_str** (*str*) – The json string to be processed.
- **citation** (*Optional[str]*) – A PubMed ID passed to be used in the evidence for the extracted INDRA Statements. Default: None

**Returns** **rp** – A ReachProcessor containing the extracted INDRA Statements in `rp.statements`.

**Return type** *ReachProcessor*

`indra.sources.reach.api.process_nxml_file` (*file\_name*, *citation=None*, *offline=False*, *output\_fname='reach\_output.json'*)

Return a ReachProcessor by processing the given NXML file.

NXML is the format used by PubmedCentral for papers in the open access subset.

##### Parameters

- **file\_name** (*str*) – The name of the NXML file to be processed.
- **citation** (*Optional[str]*) – A PubMed ID passed to be used in the evidence for the extracted INDRA Statements. Default: None
- **offline** (*Optional[bool]*) – If set to True, the REACH system is ran offline. Otherwise (by default) the web service is called. Default: False
- **output\_fname** (*Optional[str]*) – The file to output the REACH JSON output to. Defaults to reach\_output.json in current working directory.

**Returns** **rp** – A ReachProcessor containing the extracted INDRA Statements in rp.statements.

**Return type** *ReachProcessor*

```
indra.sources.reach.api.process_nxml_str(nxml_str, citation=None, offline=False, output_fname='reach_output.json')
```

Return a ReachProcessor by processing the given NXML string.

NXML is the format used by PubmedCentral for papers in the open access subset.

#### Parameters

- **nxml\_str** (*str*) – The NXML string to be processed.
- **citation** (*Optional[str]*) – A PubMed ID passed to be used in the evidence for the extracted INDRA Statements. Default: None
- **offline** (*Optional[bool]*) – If set to True, the REACH system is ran offline. Otherwise (by default) the web service is called. Default: False
- **output\_fname** (*Optional[str]*) – The file to output the REACH JSON output to. Defaults to reach\_output.json in current working directory.

**Returns** **rp** – A ReachProcessor containing the extracted INDRA Statements in rp.statements.

**Return type** *ReachProcessor*

```
indra.sources.reach.api.process_pmc(pmc_id, offline=False, output_fname='reach_output.json')
```

Return a ReachProcessor by processing a paper with a given PMC id.

Uses the PMC client to obtain the full text. If it's not available, None is returned.

#### Parameters

- **pmc\_id** (*str*) – The ID of a PubmedCentral article. The string may start with PMC but passing just the ID also works. Examples: 3717945, PMC3717945 <https://www.ncbi.nlm.nih.gov/pmc/>
- **offline** (*Optional[bool]*) – If set to True, the REACH system is ran offline. Otherwise (by default) the web service is called. Default: False

**Returns** **rp** – A ReachProcessor containing the extracted INDRA Statements in rp.statements.

**Return type** *ReachProcessor*

```
indra.sources.reach.api.process_pubmed_abstract(pubmed_id, offline=False, output_fname='reach_output.json', **kwargs)
```

Return a ReachProcessor by processing an abstract with a given Pubmed id.

Uses the Pubmed client to get the abstract. If that fails, None is returned.

#### Parameters

- **pubmed\_id** (*str*) – The ID of a Pubmed article. The string may start with PMID but passing just the ID also works. Examples: 27168024, PMID27168024 <https://www.ncbi.nlm.nih.gov/pubmed/>
- **offline** (*Optional[bool]*) – If set to True, the REACH system is ran offline. Otherwise (by default) the web service is called. Default: False
- **output\_fname** (*Optional[str]*) – The file to output the REACH JSON output to. Defaults to reach\_output.json in current working directory.
- **\*\*kwargs** (*keyword arguments*) – All other keyword arguments are passed directly to *process\_text*.

**Returns** **rp** – A ReachProcessor containing the extracted INDRA Statements in rp.statements.

**Return type** *ReachProcessor*

```
indra.sources.reach.api.process_text(text, citation=None, offline=False, output_fname='reach_output.json', timeout=None)
```

Return a ReachProcessor by processing the given text.

#### Parameters

- **text** (*str*) – The text to be processed.
- **citation** (*Optional[str]*) – A PubMed ID passed to be used in the evidence for the extracted INDRA Statements. This is used when the text to be processed comes from a publication that is not otherwise identified. Default: None
- **offline** (*Optional[bool]*) – If set to True, the REACH system is ran offline. Otherwise (by default) the web service is called. Default: False
- **output\_fname** (*Optional[str]*) – The file to output the REACH JSON output to. Defaults to reach\_output.json in current working directory.
- **timeout** (*Optional[float]*) – This only applies when reading online (*offline=False*). Only wait for *timeout* seconds for the api to respond.

**Returns** **rp** – A ReachProcessor containing the extracted INDRA Statements in rp.statements.

**Return type** *ReachProcessor*

## REACH Processor (`indra.sources.reach.processor`)

```
class indra.sources.reach.processor.ReachProcessor(json_dict, pmid=None)
```

The ReachProcessor extracts INDRA Statements from REACH parser output.

#### Parameters

- **json\_dict** (*dict*) – A JSON dictionary containing the REACH extractions.
- **pmid** (*Optional[str]*) – The PubMed ID associated with the extractions. This can be passed in case the PMID cannot be determined from the extractions alone.

#### tree

The objectpath Tree object representing the extractions.

**Type** objectpath.Tree

#### statements

A list of INDRA Statements that were extracted by the processor.

**Type** list[indra.statements.Statement]

**citation**

The PubMed ID associated with the extractions.

**Type** str

**all\_events**

The frame IDs of all events by type in the REACH extraction.

**Type** dict[str, str]

**get\_activation()**

Extract INDRA Activation Statements.

**get\_all\_events()**

Gather all event IDs in the REACH output by type.

These IDs are stored in the self.all\_events dict.

**get\_complexes()**

Extract INDRA Complex Statements.

**get\_modifications()**

Extract Modification INDRA Statements.

**get\_regulate\_amounts()**

Extract RegulateAmount INDRA Statements.

**get\_translocation()**

Extract INDRA Translocation Statements.

**print\_event\_statistics()**

Print the number of events in the REACH output by type.

**class** indra.sources.reach.processor.**Site** (*residue, position*)

**position**

Alias for field number 1

**residue**

Alias for field number 0

indra.sources.reach.processor.**determine\_reach\_subtype** (*event\_name*)

Returns the category of reach rule from the reach rule instance.

Looks at a list of regular expressions corresponding to reach rule types, and returns the longest regexp that matches, or None if none of them match.

**Parameters** **evidence** (*indra.statements.Evidence*) – A reach evidence object to subtype

**Returns** **best\_match** – A regular expression corresponding to the reach rule that was used to extract this evidence

**Return type** str

**REACH reader** (*indra.sources.reach.reader*)

**exception** indra.sources.reach.reader.**ReachOfflineReadingError**

**class** indra.sources.reach.reader.**ReachReader**

The ReachReader wraps a singleton instance of the REACH reader.

This allows calling the reader many times without having to wait for it to start up each time.

**api\_ruler**

An instance of the REACH ApiRuler class (java object).

**Type** org.clulab.reach.apis.ApiRuler

**get\_api\_ruler()**

Return the existing reader if it exists or launch a new one.

**Returns** **api\_ruler** – An instance of the REACH ApiRuler class (java object).

**Return type** org.clulab.reach.apis.ApiRuler

**TRIPS (indra.sources.trips)****TRIPS API (indra.sources.trips.api)**

indra.sources.trips.api.**process\_text** (*text*, *save\_xml\_name='trips\_output.xml'*,  
*save\_xml\_pretty=True*, *offline=False*, *service\_endpoint='drum'*)

Return a TripsProcessor by processing text.

**Parameters**

- **text** (*str*) – The text to be processed.
- **save\_xml\_name** (*Optional[str]*) – The name of the file to save the returned TRIPS extraction knowledge base XML. Default: trips\_output.xml
- **save\_xml\_pretty** (*Optional[bool]*) – If True, the saved XML is pretty-printed. Some third-party tools require non-pretty-printed XMLs which can be obtained by setting this to False. Default: True
- **offline** (*Optional[bool]*) – If True, offline reading is used with a local instance of DRUM, if available. Default: False
- **service\_endpoint** (*Optional[str]*) – Selects the TRIPS/DRUM web service endpoint to use. Is a choice between “drum” (default) and “drum-dev”, a nightly build.

**Returns** **tp** – A TripsProcessor containing the extracted INDRA Statements in tp.statements.

**Return type** *TripsProcessor*

indra.sources.trips.api.**process\_xml** (*xml\_string*)

Return a TripsProcessor by processing a TRIPS EKB XML string.

**Parameters** **xml\_string** (*str*) – A TRIPS extraction knowledge base (EKB) string to be processed. <http://trips.ihmc.us/parser/api.html>

**Returns** **tp** – A TripsProcessor containing the extracted INDRA Statements in tp.statements.

**Return type** *TripsProcessor*

indra.sources.trips.api.**process\_xml\_file** (*file\_name*)

Return a TripsProcessor by processing a TRIPS EKB XML file.

**Parameters** **file\_name** (*str*) – Path to a TRIPS extraction knowledge base (EKB) file to be processed.

**Returns** **tp** – A TripsProcessor containing the extracted INDRA Statements in tp.statements.

**Return type** *TripsProcessor*

**TRIPS Processor** (`indra.sources.trips.processor`)

**class** `indra.sources.trips.processor.TripsProcessor` (*xml\_string*)

The TripsProcessor extracts INDRA Statements from a TRIPS XML.

For more details on the TRIPS EKB XML format, see <http://trips.ihmc.us/parser/cgi/drum>

**Parameters** `xml_string` (*str*) – A TRIPS extraction knowledge base (EKB) in XML format as a string.

**tree**

An ElementTree object representation of the TRIPS EKB XML.

**Type** `xml.etree.ElementTree.Element`

**statements**

A list of INDRA Statements that were extracted from the EKB.

**Type** `list[indra.statements.Statement]`

**doc\_id**

The PubMed ID of the paper that the extractions are from.

**Type** `str`

**sentences**

The list of all sentences in the EKB with their IDs

**Type** `dict[str: str]`

**paragraphs**

The list of all paragraphs in the EKB with their IDs

**Type** `dict[str: str]`

**par\_to\_sec**

A map from paragraph IDs to their associated section types

**Type** `dict[str: str]`

**extracted\_events**

A list of Event elements that have been extracted as INDRA Statements.

**Type** `list[xml.etree.ElementTree.Element]`

**get\_activations** ()

Extract direct Activation INDRA Statements.

**get\_activations\_causal** ()

Extract causal Activation INDRA Statements.

**get\_activations\_stimulate** ()

Extract Activation INDRA Statements via stimulation.

**get\_active\_forms** ()

Extract ActiveForm INDRA Statements.

**get\_active\_forms\_state** ()

Extract ActiveForm INDRA Statements.

**get\_agents** ()

Return list of INDRA Agents corresponding to TERMS in the EKB.

This is meant to be used when entities e.g. “phosphorylated ERK”, rather than events need to be extracted from processed natural language. These entities with their respective states are represented as INDRA Agents.

**Returns agents** – List of INDRA Agents extracted from EKB.

**Return type** list[indra.statements.Agent]

**get\_all\_events** ()

Make a list of all events in the TRIPS EKB.

The events are stored in self.all\_events.

**get\_complexes** ()

Extract Complex INDRA Statements.

**get\_degradations** ()

Extract Degradation INDRA Statements.

**get\_modifications** ()

Extract all types of Modification INDRA Statements.

**get\_modifications\_indirect** ()

Extract indirect Modification INDRA Statements.

**get\_regulate\_amounts** ()

Extract Increase/DecreaseAmount Statements.

**get\_syntheses** ()

Extract IncreaseAmount INDRA Statements.

**get\_term\_agents** ()

Return dict of INDRA Agents keyed by corresponding TERMS in the EKB.

This is meant to be used when entities e.g. “phosphorylated ERK”, rather than events need to be extracted from processed natural language. These entities with their respective states are represented as INDRA Agents. Further, each key of the dictionary corresponds to the ID assigned by TRIPS to the given TERM that the Agent was extracted from.

**Returns agents** – Dict of INDRA Agents extracted from EKB.

**Return type** dict[str, indra.statements.Agent]

## TRIPS Web-service Client (`indra.sources.trips.client`)

`indra.sources.trips.client.get_xml` (*html*, *content\_tag='ekb'*, *fail\_if\_empty=False*)

Extract the content XML from the HTML output of the TRIPS web service.

### Parameters

- **html** (*str*) – The HTML output from the TRIPS web service.
- **content\_tag** (*str*) – The xml tag used to label the content. Default is ‘ekb’.
- **fail\_if\_empty** (*bool*) – If True, and if the xml content found is an empty string, raise an exception. Default is False.

### Returns

- *The extraction knowledge base (e.g. EKB) XML that contains the event and*
- *term extractions.*



`indra.sources.trips.client.save_xml(xml_str, file_name, pretty=True)`  
 Save the TRIPS EKB XML in a file.

#### Parameters

- **xml\_str** (*str*) – The TRIPS EKB XML string to be saved.
- **file\_name** (*str*) – The name of the file to save the result in.
- **pretty** (*Optional[bool]*) – If True, the XML is pretty printed.

`indra.sources.trips.client.send_query(text, service_endpoint='drum', query_args=None)`  
 Send a query to the TRIPS web service.

#### Parameters

- **text** (*str*) – The text to be processed.
- **service\_endpoint** (*Optional[str]*) – Selects the TRIPS/DRUM web service endpoint to use. Is a choice between “drum” (default), “drum-dev”, a nightly build, and “cwms” for use with more general knowledge extraction.
- **query\_args** (*Optional[dict]*) – A dictionary of arguments to be passed with the query.

**Returns** `html` – The HTML result returned by the web service.

**Return type** `str`

### TRIPS/DRUM Local Reader (`indra.sources.trips.drum_reader`)

**class** `indra.sources.trips.drum_reader.DrumReader` (*\*\*kwargs*)  
 Agent which processes text through a local TRIPS/DRUM instance.

This class is implemented as a communicative agent which sends and receives KQML messages through a socket. It sends text (ideally in small blocks like one sentence at a time) to the running DRUM instance and receives extraction knowledge base (EKB) XML responses asynchronously through the socket. To install DRUM and its dependencies locally, follow instructions at: <https://github.com/wdebeaum/drum> Once installed, run `drum/bin/trips-drum -nouser` to run DRUM without a GUI. Once DRUM is running, this class can be instantiated as `dr = DrumReader()`, at which point it attempts to connect to DRUM via the socket. You can use `dr.read_text(text)` to send text for reading. In another usage more, `dr.read_pmc(pmcid)` can be used to read a full open-access PMC paper. Receiving responses can be started as `dr.start()` which waits for responses from the reader and returns when all responses were received. Once finished, the list of EKB XML extractions can be accessed via `dr.extractions`.

#### Parameters

- **run\_drum** (*Optional[bool]*) – If True, the DRUM reading system is launched as a subprocess for reading. If False, DRUM is expected to be running independently. Default: False
- **drum\_system** (*Optional[subprocess.Popen]*) – A handle to the subprocess of a running DRUM system instance. This can be passed in in case the instance is to be reused rather than restarted. Default: None
- **\*\*kwargs** – All other keyword arguments are passed through to the DrumReader KQML module’s constructor.

#### **extractions**

A list of EKB XML extractions corresponding to the input text list.

**Type** `list[str]`

**drum\_system**

A subprocess handle that points to a running instance of the DRUM reading system. In case the DRUM system is running independently, this is None.

**Type** subprocess.Popen

**read\_pmc** (*pmcid*)

Read a given PMC article.

**Parameters** **pmcid** (*str*) – The PMC ID of the article to read. Note that only articles in the open-access subset of PMC will work.

**read\_text** (*text*)

Read a given text phrase.

**Parameters** **text** (*str*) – The text to read. Typically a sentence or a paragraph.

**receive\_reply** (*msg, content*)

Handle replies with reading results.

**Sparser** (`indra.sources.sparser`)**Sparser API** (`indra.sources.sparser.api`)

Provides an API used to run and get Statements from the Sparser reading system.

`indra.sources.sparser.api.process_text` (*text*, *output\_fmt='json'*, *outbuf=None*,  
*cleanup=True, key=""*, *\*\*kwargs*)

Return processor with Statements extracted by reading text with Sparser.

**Parameters**

- **text** (*str*) – The text to be processed
- **output\_fmt** (*Optional[str]*) – The output format to obtain from Sparser, with the two options being 'json' and 'xml'. Default: 'json'
- **outbuf** (*Optional[file]*) – A file like object that the Sparser output is written to.
- **cleanup** (*Optional[bool]*) – If True, the temporary file created, which is used as an input file for Sparser, as well as the output file created by Sparser are removed. Default: True
- **key** (*Optional[str]*) – A key which is embedded into the name of the temporary file passed to Sparser for reading. Default is empty string.

**Returns**

- *SparserXMLProcessor* or *SparserJSONProcessor* depending on what output
- *format was chosen.*

`indra.sources.sparser.api.process_nxml_str` (*nxml\_str*, *output\_fmt='json'*, *outbuf=None*,  
*cleanup=True, key=""*, *\*\*kwargs*)

Return processor with Statements extracted by reading an NXML string.

**Parameters**

- **nxml\_str** (*str*) – The string value of the NXML-formatted paper to be read.
- **output\_fmt** (*Optional[str]*) – The output format to obtain from Sparser, with the two options being 'json' and 'xml'. Default: 'json'
- **outbuf** (*Optional[file]*) – A file like object that the Sparser output is written to.

- **cleanup** (*Optional[bool]*) – If True, the temporary file created in this function, which is used as an input file for Sparser, as well as the output file created by Sparser are removed. Default: True
- **key** (*Optional[str]*) – A key which is embedded into the name of the temporary file passed to Sparser for reading. Default is empty string.

#### Returns

- *SparserXMLProcessor or SparserJSONProcessor depending on what output format was chosen.*

`indra.sources.sparser.api.process_nxml_file` (*fname*, *output\_fmt='json'*, *outbuf=None*, *cleanup=True*, *\*\*kwargs*)

Return processor with Statements extracted by reading an NXML file.

#### Parameters

- **fname** (*str*) – The path to the NXML file to be read.
- **output\_fmt** (*Optional[str]*) – The output format to obtain from Sparser, with the two options being 'json' and 'xml'. Default: 'json'
- **outbuf** (*Optional[file]*) – A file like object that the Sparser output is written to.
- **cleanup** (*Optional[bool]*) – If True, the output file created by Sparser is removed. Default: True

#### Returns

- *sp (SparserXMLProcessor or SparserJSONProcessor depending on what output format was chosen.)*

`indra.sources.sparser.api.process_sparser_output` (*output\_fname*, *output\_fmt='json'*)

Return a processor with Statements extracted from Sparser XML or JSON

#### Parameters

- **output\_fname** (*str*) – The path to the Sparser output file to be processed. The file can either be JSON or XML output from Sparser, with the `output_fmt` parameter defining what format is assumed to be processed.
- **output\_fmt** (*Optional[str]*) – The format of the Sparser output to be processed, can either be 'json' or 'xml'. Default: 'json'

#### Returns

- *sp (SparserXMLProcessor or SparserJSONProcessor depending on what output format was chosen.)*

`indra.sources.sparser.api.process_json_dict` (*json\_dict*)

Return processor with Statements extracted from a Sparser JSON.

**Parameters** `json_dict` (*dict*) – The JSON object obtained by reading content with Sparser, using the 'json' output mode.

**Returns** `sp` – A SparserJSONProcessor which has extracted Statements as its statements attribute.

**Return type** SparserJSONProcessor

`indra.sources.sparser.api.process_xml` (*xml\_str*)

Return processor with Statements extracted from a Sparser XML.

**Parameters** `xml_str` (*str*) – The XML string obtained by reading content with Sparser, using the ‘xml’ output mode.

**Returns** `sp` – A SparserXMLProcessor which has extracted Statements as its statements attribute.

**Return type** SparserXMLProcessor

`indra.sources.sparser.api.run_sparser` (*fname, output\_fmt, outbuf=None, timeout=600*)

Return the path to reading output after running Sparser reading.

**Parameters**

- **fname** (*str*) – The path to an input file to be processed. Due to the Spaser executable’s assumptions, the file name needs to start with PMC and should be an NXML formatted file.
- **output\_fmt** (*Optional[str]*) – The format in which Sparser should produce its output, can either be ‘json’ or ‘xml’.
- **outbuf** (*Optional[file]*) – A file like object that the Sparser output is written to.
- **timeout** (*int*) – The number of seconds to wait until giving up on this one reading. The default is 600 seconds (i.e. 10 minutes). Sparser is a fast reader and the typical type to read a single full text is a matter of seconds.

**Returns** `output_path` – The path to the output file created by Sparser.

**Return type** `str`

`indra.sources.sparser.api.get_version` ()

Return the version of the Sparser executable on the path.

**Returns** `version` – The version of Sparser that is found on the Sparser path.

**Return type** `str`

`indra.sources.sparser.api.make_nxml_from_text` (*text*)

Return raw text wrapped in NXML structure.

**Parameters** `text` (*str*) – The raw text content to be wrapped in an NXML structure.

**Returns** `nxml_str` – The NXML string wrapping the raw text input.

**Return type** `str`

## MedScan (`indra.sources.medscan`)

MedScan is Elsevier’s proprietary text-mining system for reading the biological literature. This INDRA module enables processing output files (in CSXML format) from the MedScan system into INDRA Statements.

### MedScan API (`indra.sources.medscan.api`)

`indra.sources.medscan.api.process_directory` (*directory\_name, lazy=False*)

Processes a directory filled with CSXML files, first normalizing the character encodings to utf-8, and then processing into a list of INDRA statements.

**Parameters**

- **directory\_name** (*str*) – The name of a directory filled with csxml files to process
- **lazy** (*bool*) – If True, the statements will not be generated immediately, but rather a generator will be formulated, and statements can be retrieved by using `iter_statements`. If False, the `statements` attribute will be populated immediately. Default is False.

**Returns mp** – A MedscanProcessor populated with INDRA statements extracted from the csxml files

**Return type** *indra.sources.medscan.processor.MedscanProcessor*

`indra.sources.medscan.api.process_directory_statements_sorted_by_pmid(directory_name)`  
Processes a directory filled with CSXML files, first normalizing the character encoding to utf-8, and then processing into INDRA statements sorted by pmid.

**Parameters directory\_name** (*str*) – The name of a directory filled with csxml files to process

**Returns pmid\_dict** – A dictionary mapping pmids to a list of statements corresponding to that pmid

**Return type** dict

`indra.sources.medscan.api.process_file(filename, interval=None, lazy=False)`  
Process a CSXML file for its relevant information.

Consider running the `fix_csxml_character_encoding.py` script in `indra/sources/medscan` to fix any encoding issues in the input file before processing.

`indra.sources.medscan.api.filename`  
The csxml file, containing Medscan XML, to process

**Type** str

`indra.sources.medscan.api.interval`  
Select the interval of documents to read, starting with the ‘start’th document and ending before the ‘end’th document. If either is None, the value is considered undefined. If the value exceeds the bounds of available documents, it will simply be ignored.

**Type** (start, end) or None

`indra.sources.medscan.api.lazy`  
If True, the statements will not be generated immediately, but rather a generator will be formulated, and statements can be retrieved by using `iter_statements`. If False, the `statements` attribute will be populated immediately. Default is False.

**Type** bool

**Returns mp** – A MedscanProcessor object containing extracted statements

**Return type** *MedscanProcessor*

`indra.sources.medscan.api.process_file_sorted_by_pmid(file_name)`  
Processes a file and returns a dictionary mapping pmids to a list of statements corresponding to that pmid.

**Parameters file\_name** (*str*) – A csxml file to process

**Returns s\_dict** – Dictionary mapping pmids to a list of statements corresponding to that pmid

**Return type** dict

## MedScan Processor (`indra.sources.medscan.processor`)

`class indra.sources.medscan.processor.MedscanEntity(name, urn, type, properties, ch_start, ch_end)`

`ch_end`

Alias for field number 5

**ch\_start**

Alias for field number 4

**name**

Alias for field number 0

**properties**

Alias for field number 3

**type**

Alias for field number 2

**urn**

Alias for field number 1

**class** `indra.sources.medscan.processor.MedscanProcessor`

Processes Medscan data into INDRA statements.

The special StateEffect event conveys information about the binding site of a protein modification. Sometimes this is paired with additional event information in a separate SVO. When we encounter a StateEffect, we don't process into an INDRA statement right away, but instead store the site information and use it if we encounter a ProtModification event within the same sentence.

**statements**

A list of extracted INDRA statements

**Type** `list<str>`

**sentence\_statements**

A list of statements for the sentence we are currently processing. Deduplicated and added to the main statement list when we finish processing a sentence.

**Type** `list<str>`

**num\_entities**

The total number of subject or object entities the processor attempted to resolve

**Type** `int`

**num\_entities\_not\_found**

The number of subject or object IDs which could not be resolved by looking in the list of entities or tagged phrases.

**Type** `int`

**last\_site\_info\_in\_sentence**

Stored protein site info from the last StateEffect event within the sentence, allowing us to combine information from StateEffect and ProtModification events within a single sentence in a single INDRA statement. This is reset at the end of each sentence

**Type** `SiteInfo`

**agent\_from\_entity** (*relation, entity\_id*)

Create a (potentially grounded) INDRA Agent object from a given Medscan entity describing the subject or object.

Uses helper functions to convert a Medscan URN to an INDRA db\_refs grounding dictionary.

If the entity has properties indicating that it is a protein with a mutation or modification, then constructs the needed ModCondition or MutCondition.

**Parameters**

- **relation** (`MedscanRelation`) – The current relation being processed

- **entity\_id** (*str*) – The ID of the entity to process

**Returns agent** – A potentially grounded INDRA agent representing this entity

**Return type** `indra.statements.Agent`

**process\_csxml\_file** (*filename, interval=None, lazy=False*)

Processes a filehandle to MedScan csxml input into INDRA statements.

The CSXML format consists of a top-level `<batch>` root element containing a series of `<doc>` (document) elements, in turn containing `<sec>` (section) elements, and in turn containing `<sent>` (sentence) elements.

Within the `<sent>` element, a series of additional elements appear in the following order:

- `<toks>`, which contains a tokenized form of the sentence in its text attribute
- `<textmods>`, which describes any preprocessing/normalization done to the underlying text
- `<match>` elements, each of which contains one or more `<entity>` elements, describing entities in the text with their identifiers. The local IDs of each entities are given in the `msid` attribute of this element; these IDs are then referenced in any subsequent SVO elements.
- `<svo>` elements, representing subject-verb-object triples. SVO elements with a `type` attribute of `CONTROL` represent normalized regulation relationships; they often represent the normalized extraction of the immediately preceding (but unnormalized SVO element). However, in some cases there can be a “CONTROL” SVO element without its parent immediately preceding it.

#### Parameters

- **filename** (*string*) – The path to a Medscan csxml file.
- **interval** (*(start, end) or None*) – Select the interval of documents to read, starting with the ‘start’th document and ending before the ‘end’th document. If either is None, the value is considered undefined. If the value exceeds the bounds of available documents, it will simply be ignored.
- **lazy** (*bool*) – If True, only create a generator which can be used by the `get_statements` method. If True, populate the statements list now.

**process\_relation** (*relation, last\_relation*)

Process a relation into an INDRA statement.

#### Parameters

- **relation** (`MedscanRelation`) – The relation to process (a CONTROL svo with normalized verb)
- **last\_relation** (`MedscanRelation`) – The relation immediately preceding the relation to process within the same sentence, or None if there are no preceding relations within the same sentence. This preceding relation, if available, will refer to the same interaction but with an unnormalized (potentially more specific) verb, and is used when processing protein modification events.

**class** `indra.sources.medscan.processor.MedscanProperty` (*type, name, urn*)

**name**

Alias for field number 1

**type**

Alias for field number 0

**urn**

Alias for field number 2

```
class indra.sources.medscan.processor.MedscanRelation(pmid, uri, sec, entities,  
tagged_sentence, subj, verb,  
obj, svo_type)
```

A structure representing the information contained in a Medscan SVO xml element as well as associated entities and properties.

**pmid**

The URI of the current document (such as a PMID)

**Type** str

**sec**

The section of the document the relation occurs in

**Type** str

**entities**

A dictionary mapping entity IDs from the same sentence to MedscanEntity objects.

**Type** dict

**tagged\_sentence**

The sentence from which the relation was extracted, with some tagged phrases and annotations.

**Type** str

**subj**

The entity ID of the subject

**Type** str

**verb**

The verb in the relationship between the subject and the object

**Type** str

**obj**

The entity ID of the object

**Type** str

**svo\_type**

The type of SVO relationship (for example, CONTROL indicates that the verb is normalized)

**Type** str

```
class indra.sources.medscan.processor.ProteinSiteInfo(site_text, object_text)
```

Represent a site on a protein, extracted from a StateEffect event.

**Parameters**

- **site\_text** (*str*) – The site as a string (ex. S22)
- **object\_text** (*str*) – The protein being modified, as the string that appeared in the original sentence

**get\_sites** ()

Parse the site-text string and return a list of sites.

**Returns** **sites** – A list of position-residue pairs corresponding to the site-text

**Return type** list[*Site*]



`indra.sources.medscan.processor.normalize_medscan_name` (*name*)

Removes the “complex” and “complex complex” suffixes from a medscan agent name so that it better corresponds with the grounding map.

**Parameters** `name` (*str*) – The Medscan agent name

**Returns** `norm_name` – The Medscan agent name with the “complex” and “complex complex” suffixes removed.

**Return type** `str`

## TEES (`indra.sources.tees`)

The TEES processor requires an installation of TEES. To install TEES:

1. Clone the latest stable version of TEES using

```
git clone https://github.com/jbjorne/TEES.git
```

2. Put this TEES cloned repository in one of these three places: the same directory as INDRA, your home directory, or ~/Downloads. If you put TEES in a location other than one of these three places, you will need to pass this directory to `indra.sources.tees.api.process_text` each time you call it.
3. Run `configure.py` within the TEES installation to install TEES dependencies.

## TEES API (`indra.sources.tees.api`)

This module provides a simplified API for invoking the Turku Event Extraction System (TEES) on text and extracting INDRA statement from TEES output.

See publication: Jari Björne, Sofie Van Landeghem, Sampo Pyysalo, Tomoko Ohta, Filip Ginter, Yves Van de Peer, Sofia Ananiadou and Tapio Salakoski, PubMed-Scale Event Extraction for Post-Translational Modifications, Epigenetics and Protein Structural Relations. Proceedings of BioNLP 2012, pages 82-90, 2012.

`indra.sources.tees.api.run_on_text` (*text*, *python2\_path*)

Runs TEES on the given text in a temporary directory and returns a temporary directory with TEES output.

The caller should delete this directory when done with it. This function runs TEES and produces TEES output files but does not process TEES output into INDRA statements.

### Parameters

- `text` (*str*) – Text from which to extract relationships
- `python2_path` (*str*) – The path to the python 2 interpreter

**Returns** `output_dir` – Temporary directory with TEES output. The caller should delete this directory when done with it.

**Return type** `str`

`indra.sources.tees.api.process_text` (*text*, *pmid=None*, *python2\_path=None*)

Processes the specified plain text with TEES and converts output to supported INDRA statements. Check for the TEES installation is the `TEES_PATH` environment variable, and configuration file; if not found, checks candidate paths in `tees_candidate_paths`. Raises an exception if TEES cannot be found in any of these places.

### Parameters

- `text` (*str*) – Plain text to process with TEES
- `pmid` (*str*) – The PMID from which the paper comes from, to be stored in the Evidence object of statements. Set to None if this is unspecified.

- **python2\_path** (*str*) – TEES is only compatible with python 2. This processor invokes this external python 2 interpreter so that the processor can be run in either python 2 or python 3. If None, searches for an executable named python2 in the PATH environment variable.

**Returns** **tp** – A TEESProcessor object which contains a list of INDRA statements extracted from TEES extractions

**Return type** *TEESProcessor*

`indra.sources.tees.api.extract_output` (*output\_dir*)

Extract the text of the a1, a2, and sentence segmentation files from the TEES output directory. These files are located within a compressed archive.

**Parameters** **output\_dir** (*str*) – Directory containing the output of the TEES system

**Returns**

- **a1\_text** (*str*) – The text of the TEES a1 file (specifying the entities)
- **a2\_text** (*str*) – The text of the TEES a2 file (specifying the event graph)
- **sentence\_segmentations** (*str*) – The text of the XML file specifying the sentence segmentation

### TEES Processor (`indra.sources.tees.processor`)

This module takes the TEES parse graph generated by `parse_tees` and converts it into INDRA statements.

See publication: Jari Björne, Sofie Van Landeghem, Sampo Pyysalo, Tomoko Ohta, Filip Ginter, Yves Van de Peer, Sofia Ananiadou and Tapio Salakoski, PubMed-Scale Event Extraction for Post-Translational Modifications, Epigenetics and Protein Structural Relations. Proceedings of BioNLP 2012, pages 82-90, 2012.

**class** `indra.sources.tees.processor.TEESProcessor` (*a1\_text*, *a2\_text*, *sentence\_segmentations*, *pmid*)

Converts the output of the TEES reader to INDRA statements.

Only extracts a subset of INDRA statements. Currently supported statements are: \* Phosphorylation \* Dephosphorylation \* Binding \* IncreaseAmount \* DecreaseAmount

**Parameters**

- **a1\_text** (*str*) – The TEES a1 output file, with entity information
- **a2\_text** (*str*) – The TEES a2 output file, with the event graph
- **sentence\_segmentations** (*str*) – The TEES sentence segmentation XML output
- **pmid** (*int*) – The pmid which the text comes from, or None if we don't want to specify at the moment. Stored in the Evidence object for each statement.

**statements**

A list of INDRA statements extracted from the provided text via TEES

**Type** `list[indra.statements.Statement]`

**connected\_subgraph** (*node*)

Returns the subgraph containing the given node, its ancestors, and its descendants.

**Parameters** **node** (*str*) – We want to create the subgraph containing this node.

**Returns** **subgraph** – The subgraph containing the specified node.

**Return type** `networkx.DiGraph`

**find\_event\_parent\_with\_event\_child** (*parent\_name, child\_name*)

Finds all event nodes (is\_event node attribute is True) that are of the type parent\_name, that have a child event node with the type child\_name.

**find\_event\_with\_outgoing\_edges** (*event\_name, desired\_relations*)

Gets a list of event nodes with the specified event\_name and outgoing edges annotated with each of the specified relations.

**Parameters**

- **event\_name** (*str*) – Look for event nodes with this name
- **desired\_relations** (*list[str]*) – Look for event nodes with outgoing edges annotated with each of these relations

**Returns** *event\_nodes* – Event nodes that fit the desired criteria

**Return type** list[str]

**general\_node\_label** (*node*)

Used for debugging - gives a short text description of a graph node.

**get\_entity\_text\_for\_relation** (*node, relation*)

Looks for an edge from node to some other node, such that the edge is annotated with the given relation. If there exists such an edge, and the node at the other edge is an entity, return that entity's text. Otherwise, returns None.

**get\_related\_node** (*node, relation*)

Looks for an edge from node to some other node, such that the edge is annotated with the given relation. If there exists such an edge, returns the name of the node it points to. Otherwise, returns None.

**node\_has\_edge\_with\_label** (*node\_name, edge\_label*)

Looks for an edge from node\_name to some other node with the specified label. Returns the node to which this edge points if it exists, or None if it doesn't.

**Parameters**

- **G** – The graph object
- **node\_name** – Node that the edge starts at
- **edge\_label** – The text in the relation property of the edge

**node\_to\_evidence** (*entity\_node, is\_direct*)

Computes an evidence object for a statement.

We assume that the entire event happens within a single statement, and get the text of the sentence by getting the text of the sentence containing the provided node that corresponds to one of the entities participating in the event.

The Evidence's pmid is whatever was provided to the constructor (perhaps None), and the annotations are the subgraph containing the provided node, its ancestors, and its descendants.

**print\_parent\_and\_children\_info** (*node*)

Used for debugging - prints a short description of a node, its children, its parents, and its parents' children.

**process\_binding\_statements** ()

Looks for Binding events in the graph and extracts them into INDRA statements.

In particular, looks for a Binding event node with outgoing edges with relations Theme and Theme2 - the entities these edges point to are the two constituents of the Complex INDRA statement.

**process\_decrease\_expression\_amount** ()

Looks for Negative\_Regulation events with a specified Cause and a Gene\_Expression theme, and processes them into INDRA statements.

**process\_increase\_expression\_amount** ()

Looks for Positive\_Regulation events with a specified Cause and a Gene\_Expression theme, and processes them into INDRA statements.

**process\_phosphorylation\_statements** ()

Looks for Phosphorylation events in the graph and extracts them into INDRA statements.

In particular, looks for a Positive\_regulation event node with a child Phosphorylation event node.

If Positive\_regulation has an outgoing Cause edge, that's the subject If Phosphorylation has an outgoing Theme edge, that's the object If Phosphorylation has an outgoing Site edge, that's the site

```
indra.sources.tees.processor.s2a(s)
```

Makes an Agent from a string describing the agent.

**ISI (indra.sources.isi)**

This module provides an input interface and processor to the ISI reading system.

The reader is set up to run within a Docker container. For the ISI reader to run, set the Docker memory and swap space to the maximum. For processing nxml files, install the nxml2txt utility (<https://github.com/spyysalo/nxml2txt>) and set the configuration variable NXML2TXT\_PATH to its location. In addition, since the reader works with Python 2 only, make sure PYTHON2\_PATH is set in your config file or environment and points to a Python 2 executable.

**ISI API (indra.sources.isi.api)**

```
indra.sources.isi.api.process_json_file(file_path, pmid=None, extra_annotations=None,
                                       add_grounding=True)
```

Extracts statements from the given ISI output file.

**Parameters**

- **file\_path** (*str*) – The ISI output file from which to extract statements
- **pmid** (*int*) – The PMID of the document being preprocessed, or None if not specified
- **extra\_annotations** (*dict*) – Extra annotations to be added to each statement from this document (can be the empty dictionary)
- **add\_grounding** (*Optional[bool]*) – If True the extracted Statements' grounding is mapped

```
indra.sources.isi.api.process_nxml(nxml_filename, pmid=None, extra_annotations=None,
                                   cleanup=True, add_grounding=True)
```

Process an NXML file using the ISI reader

First converts NXML to plain text and preprocesses it, then runs the ISI reader, and processes the output to extract INDRA Statements.

**Parameters**

- **nxml\_filename** (*str*) – nxml file to process
- **pmid** (*Optional[str]*) – pmid of this nxml file, to be added to the Evidence object of the extracted INDRA statements

- **extra\_annotations** (*Optional[dict]*) – Additional annotations to add to the Evidence object of all extracted INDRA statements. Extra annotations called ‘interaction’ are ignored since this is used by the processor to store the corresponding raw ISI output.
- **cleanup** (*Optional[bool]*) – If True, the temporary folders created for preprocessed reading input and output are removed. Default: True
- **add\_grounding** (*Optional[bool]*) – If True the extracted Statements’ grounding is mapped

**Returns ip** – A processor containing extracted Statements

**Return type** *indra.sources.isi.processor.IsiProcessor*

```
indra.sources.isi.api.process_output_folder(folder_path, pmids=None,
                                           extra_annotations=None,
                                           add_grounding=True)
```

Recursively extracts statements from all ISI output files in the given directory and subdirectories.

#### Parameters

- **folder\_path** (*str*) – The directory to traverse
- **pmids** (*Optional[str]*) – PMID mapping to be added to the Evidence of the extracted INDRA Statements
- **extra\_annotations** (*Optional[dict]*) – Additional annotations to add to the Evidence object of all extracted INDRA statements. Extra annotations called ‘interaction’ are ignored since this is used by the processor to store the corresponding raw ISI output.
- **add\_grounding** (*Optional[bool]*) – If True the extracted Statements’ grounding is mapped

```
indra.sources.isi.api.process_preprocessed(isi_preprocessor, num_processes=1,
                                          output_dir=None, cleanup=True,
                                          add_grounding=True)
```

Process a directory of abstracts and/or papers preprocessed using the specified IsiPreprocessor, to produce a list of extracted INDRA statements.

#### Parameters

- **isi\_preprocessor** (*indra.sources.isi.preprocessor.IsiPreprocessor*) – Preprocessor object that has already preprocessed the documents we want to read and process with the ISI reader
- **num\_processes** (*Optional[int]*) – Number of processes to parallelize over
- **output\_dir** (*Optional[str]*) – The directory into which to put reader output; if omitted or None, uses a temporary directory.
- **cleanup** (*Optional[bool]*) – If True, the temporary folders created for preprocessed reading input and output are removed. Default: True
- **add\_grounding** (*Optional[bool]*) – If True the extracted Statements’ grounding is mapped

**Returns ip** – A processor containing extracted statements

**Return type** *indra.sources.isi.processor.IsiProcessor*

```
indra.sources.isi.api.process_text(text, pmid=None, cleanup=True, add_grounding=True)
```

Process a string using the ISI reader and extract INDRA statements.

#### Parameters

- **text** (*str*) – A text string to process
- **pmid** (*Optional[str]*) – The PMID associated with this text (or None if not specified)
- **cleanup** (*Optional[bool]*) – If True, the temporary folders created for preprocessed reading input and output are removed. Default: True
- **add\_grounding** (*Optional[bool]*) – If True the extracted Statements’ grounding is mapped

**Returns ip** – A processor containing statements

**Return type** *indra.sources.isi.processor.IsiProcessor*

### ISI Processor (*indra.sources.isi.processor*)

```
class indra.sources.isi.processor.IsiProcessor (reader_output,          pmid=None,  
                                              extra_annotations=None,  
                                              add_grounding=True)
```

Processes the output of the ISI reader.

**reader\_output**

The output JSON of the ISI reader as a json object.

**Type** json

**verbs**

A list of verbs that have appeared in the processed ISI output

**Type** set[str]

**pmid**

The PMID to assign to the extracted Statements

**Type** str

**extra\_annotations**

Annotations to be included with each extracted Statement

**Type** dict

**statements**

Extracted statements

**Type** list[indra.statements.Statement]

**get\_statements** ()

Process reader output to produce INDRA Statements.

### Geneways (*indra.sources.geneways*)

#### Geneways API (*indra.sources.geneways.api*)

This module provides a simplified API for invoking the Geneways input processor , which converts extracted information collected with Geneways into INDRA statements.

See publication: Rzhetsky, Andrey, Ivan Iossifov, Tomohiro Koike, Michael Krauthammer, Pauline Kra, Mitzi Morris, Hong Yu et al. “GeneWays: a system for extracting, analyzing, visualizing, and integrating molecular pathway data.” Journal of biomedical informatics 37, no. 1 (2004): 43-53.

`indra.sources.geneways.api.process_geneways_files` (*input\_folder*='/home/docs/checkouts/readthedocs.org/user\_  
get\_evidence=True)

Reads in Geneways data and returns a list of statements.

#### Parameters

- **input\_folder** (*Optional[str]*) – A folder in which to search for Geneways data. Looks for these Geneways extraction data files: `human_action.txt`, `human_actionmention.txt`, `human_symbols.txt`. Omit this parameter to use the default input folder which is `indra/data`.
- **get\_evidence** (*Optional[bool]*) – Attempt to find the evidence text for an extraction by downloading the corresponding text content and searching for the given offset in the text to get the evidence sentence. Default: True

**Returns** `gp` – A `GenewaysProcessor` object which contains a list of INDRA statements generated from the Geneways action mentions.

**Return type** *GenewaysProcessor*

### Geneways Processor (`indra.sources.geneways.processor`)

This module provides an input processor for information extracted using the Geneways software suite, converting extraction data in Geneways format into INDRA statements.

See publication: Rzhetsky, Andrey, Ivan Iossifov, Tomohiro Koike, Michael Krauthammer, Pauline Kra, Mitzi Morris, Hong Yu et al. “GeneWays: a system for extracting, analyzing, visualizing, and integrating molecular pathway data.” *Journal of biomedical informatics* 37, no. 1 (2004): 43-53.

**class** `indra.sources.geneways.processor.GenewaysProcessor` (*search\_path*,  
*get\_evidence=True*)

The `GenewaysProcessors` converts extracted Geneways action mentions into INDRA statements.

**Parameters** **search\_path** (*list[str]*) – A list of directories in which to search for Geneways data

#### statements

A list of INDRA statements converted from Geneways action mentions, populated by calling the constructor

**Type** `list[indra.statements.Statement]`

**make\_statement** (*action, mention*)

Makes an INDRA statement from a Geneways action and action mention.

#### Parameters

- **action** (*GenewaysAction*) – The mechanism that the Geneways mention maps to. Note that several text mentions can correspond to the same action if they are referring to the same relationship - there may be multiple Geneways action mentions corresponding to each action.
- **mention** (*GenewaysActionMention*) – The Geneways action mention object corresponding to a single mention of a mechanism in a specific text. We make a new INDRA statement corresponding to each action mention.

**Returns** **statement** – An INDRA statement corresponding to the provided Geneways action mention, or `None` if the action mention’s type does not map onto any INDRA statement type in `geneways_action_type_mapper`.

**Return type** `indra.statements.Statement`

`indra.sources.geneways.processor.geneways_action_to_indra_statement_type` (*actiontype*, *plo*)

Return INDRA Statement corresponding to Geneways action type.

#### Parameters

- **actiontype** (*str*) – The verb extracted by the Geneways processor
- **plo** (*str*) – A one character string designating whether Geneways classifies this verb as a physical, logical, or other interaction

**Returns** If there is no mapping to INDRA statements from this action type the return value is None. If there is such a mapping, `statement_generator` is an anonymous function that takes in the subject agent, object agent, and evidence, in that order, and returns an INDRA statement object.

**Return type** `statement_generator`

### RLIMS-P (`indra.sources.rlimsp`)

RLIMS-P is a rule-based reading system which extracts phosphorylation relationships with sites from text. RLIMS-P exposes a web service to submit PubMed IDs and PMC IDs for processing.

See also: <https://research.bioinformatics.udel.edu/rlimsp/> and <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4568560/>

### RLIMS-P API (`indra.sources.rlimsp.api`)

`indra.sources.rlimsp.api.process_from_webservice` (*id\_val*, *id\_type='pmcid'*,  
*source='pmc'*,  
*with\_grounding=True*)

Return an output from RLIMS-p for the given PubMed ID or PMC ID.

#### Parameters

- **id\_val** (*str*) – A PMCID, with the prefix PMC, or pmid, with no prefix, of the paper to be “read”.
- **id\_type** (*str*) – Either ‘pmid’ or ‘pmcid’. The default is ‘pmcid’.
- **source** (*str*) – Either ‘pmc’ or ‘medline’, whether you want pmc fulltext or medline abstracts.
- **with\_grounding** (*bool*) – The RLIMS-P web service provides two endpoints, one pre-grounded, the other not so much. The grounded endpoint returns far less content, and may perform some grounding that can be handled by the grounding mapper.

**Returns** An `RlimspProcessor` which contains a list of extracted INDRA Statements in its `statements` attribute.

**Return type** `indra.sources.rlimsp.processor.RlimspProcessor`

`indra.sources.rlimsp.api.process_from_json_file` (*filename*, *doc\_id\_type=None*)

Process RLIMSP extractions from a bulk-download JSON file.

#### Parameters

- **filename** (*str*) – Path to the JSON file.
- **doc\_id\_type** (*Optional[str]*) – In some cases the RLIMS-P paragraph info doesn’t contain ‘pmid’ or ‘pmcid’ explicitly, instead it contains a ‘docId’ key. This parameter allows



defining what ID type ‘docId’ could be interpreted as. Its values should be ‘pmid’ or ‘pmcid’ or None if not used.

**Returns** An RlimspProcessor which contains a list of extracted INDRA Statements in its statements attribute.

**Return type** `indra.sources.rlimsp.processor.RlimspProcessor`

`indra.sources.rlimsp.api.process_from_jsonish_str(jsonish_str, doc_id_type=None)`  
Process RLIMSP extractions from a bulk-download JSON file.

**Parameters**

- **jsonish\_str** (*str*) – The contents of one of the not-quite-json files you can find here: <https://hershey.dbi.udel.edu/textmining/export>
- **doc\_id\_type** (*Optional[str]*) – In some cases the RLIMS-P paragraph info doesn’t contain ‘pmid’ or ‘pmcid’ explicitly, instead it contains a ‘docId’ key. This parameter allows defining what ID type ‘docId’ could be interpreted as. Its values should be ‘pmid’ or ‘pmcid’ or None if not used.

**Returns** An RlimspProcessor which contains a list of extracted INDRA Statements in its statements attribute.

**Return type** `indra.sources.rlimsp.processor.RlimspProcessor`

### RLIMSP-P Processor (`indra.sources.rlimsp.processor`)

**class** `indra.sources.rlimsp.processor.RlimspParagraph` (*p\_info, doc\_id\_type*)

An object that represents a single RLIMS-P Paragraph.

**class** `indra.sources.rlimsp.processor.RlimspProcessor` (*rlimsp\_json, doc\_id\_type=None*)

Convert RLIMS-P JSON into INDRA Statements.

**extract\_statements** ()

Extract the statements from the json.

`indra.sources.rlimsp.processor.get_agent_from_entity_info(entity_info)`

Return an INDRA Agent by processing an entity\_info dict.

## 4.2.2 General Purpose Reading Systems

### Eidos (`indra.sources.eidos`)

Eidos is an open-domain machine reading system which uses a cascade of grammars to extract causal events from free text. It is ideal for modeling applications that are not specific to a given domain like molecular biology.

To set up reading with Eidos, the Eidos system and its dependencies need to be compiled and packaged as a fat JAR:

```
git clone https://github.com/clulab/eidos.git
cd eidos
sbt assembly
```

This creates a JAR file in `eidos/target/scala[version]/eidos-[version].jar`. Set the absolute path to this file on the `EIDOSPATH` environmental variable and then append `EIDOSPATH` to the `CLASSPATH` environmental variable (entries are separated by colons).

The *pyjnius* package needs to be set up and operational to use Eidos reading in Python. For more details, see [Pyjnius](#) setup instructions in the documentation.

For eidos to provide grounding information to be included in INDRA Statements, the eidos configuration needs to be adjusted. First, download `vectors.txt` from <https://drive.google.com/open?id=1tffQuLB5XtKcq9wlo0n-tsnPJYQF18oS> and put it in a folder called `src/main/resources/org/clulab/wm/eidos/w2v` within your eidos folder. Next, set the property “useW2V” to true in `src/main/resources/eidos.conf`. Finally, rerun sbt assembly.

### Eidos API (`indra.sources.eidos.api`)

`indra.sources.eidos.api.initialize_reader()`

Instantiate an Eidos reader for fast subsequent reading.

`indra.sources.eidos.api.process_json(json_dict, grounding_ns=None)`

Return an EidosProcessor by processing a Eidos JSON-LD dict.

#### Parameters

- **json\_dict** (*dict*) – The JSON-LD dict to be processed.
- **grounding\_ns** (*Optional[list]*) – A list of name spaces for which INDRA should represent groundings, when given. If not specified or None, all grounding name spaces are propagated. If an empty list, no groundings are propagated. Example: ['UN', 'WM'], Default: None

**Returns ep** – A EidosProcessor containing the extracted INDRA Statements in its statements attribute.

**Return type** *EidosProcessor*

`indra.sources.eidos.api.process_json_file(file_name, grounding_ns=None)`

Return an EidosProcessor by processing the given Eidos JSON-LD file.

This function is useful if the output from Eidos is saved as a file and needs to be processed.

#### Parameters

- **file\_name** (*str*) – The name of the JSON-LD file to be processed.
- **grounding\_ns** (*Optional[list]*) – A list of name spaces for which INDRA should represent groundings, when given. If not specified or None, all grounding name spaces are propagated. If an empty list, no groundings are propagated. Example: ['UN', 'WM'], Default: None

**Returns ep** – A EidosProcessor containing the extracted INDRA Statements in its statements attribute.

**Return type** *EidosProcessor*

`indra.sources.eidos.api.process_json_str(json_str, grounding_ns=None)`

Return an EidosProcessor by processing the Eidos JSON-LD string.

#### Parameters

- **json\_str** (*str*) – The JSON-LD string to be processed.
- **grounding\_ns** (*Optional[list]*) – A list of name spaces for which INDRA should represent groundings, when given. If not specified or None, all grounding name spaces are propagated. If an empty list, no groundings are propagated. Example: ['UN', 'WM'], Default: None

**Returns ep** – A EidosProcessor containing the extracted INDRA Statements in its statements attribute.

**Return type** *EidosProcessor*

```
indra.sources.eidos.api.process_text(text, out_format='json_ld',
                                     save_json='eidos_output.json', webservice=None,
                                     grounding_ns=None)
```

Return an EidosProcessor by processing the given text.

This constructs a reader object via Java and extracts mentions from the text. It then serializes the mentions into JSON and processes the result with process\_json.

#### Parameters

- **text** (*str*) – The text to be processed.
- **out\_format** (*Optional[str]*) – The type of Eidos output to read into and process. Currently only 'json-ld' is supported which is also the default value used.
- **save\_json** (*Optional[str]*) – The name of a file in which to dump the JSON output of Eidos.
- **webservice** (*Optional[str]*) – An Eidos reader web service URL to send the request to. If None, the reading is assumed to be done with the Eidos JAR rather than via a web service. Default: None
- **grounding\_ns** (*Optional[list]*) – A list of name spaces for which INDRA should represent groundings, when given. If not specified or None, all grounding name spaces are propagated. If an empty list, no groundings are propagated. Example: ['UN', 'WM'], Default: None

**Returns ep** – An EidosProcessor containing the extracted INDRA Statements in its statements attribute.

**Return type** *EidosProcessor*

### Eidos Processor (`indra.sources.eidos.processor`)

```
class indra.sources.eidos.processor.EidosProcessor(json_dict, grounding_ns=None)
```

This processor extracts INDRA Statements from Eidos JSON-LD output.

**Parameters json\_dict** (*dict*) – A JSON dictionary containing the Eidos extractions in JSON-LD format.

#### statements

A list of INDRA Statements that were extracted by the processor.

**Type** list[indra.statements.Statement]

#### extract\_causal\_relations()

Extract causal relations as Statements.

#### geo\_context\_from\_ref(ref)

Return a ref context object given a location reference entry.

#### get\_concept(entity)

Return Concept from an Eidos entity.

#### get\_evidence(relation)

Return the Evidence object for the INDRA Statement.

**get\_groundings** (*entity*)

Return groundings as db\_refs for an entity.

**static get\_hedging** (*event*)

Return hedging markers attached to an event.

**Example:** “states”: [{"@type": “State”, “type”: “HEDGE”, “text”: “could”}]

**static get\_negation** (*event*)

Return negation attached to an event.

**Example:** “states”: [{"@type": “State”, “type”: “NEGATION”, “text”: “n’t”}]

**time\_context\_from\_ref** (*timex*)

Return a time context object given a timex reference entry.

`indra.sources.eidos.processor.find_arg` (*event, arg\_type*)

Return ID of the first argument of a given type

`indra.sources.eidos.processor.find_args` (*event, arg\_type*)

Return IDs of all arguments of a given type

`indra.sources.eidos.processor.ref_context_from_geoloc` (*geoloc*)

Return a RefContext object given a geoloc entry.

`indra.sources.eidos.processor.time_context_from_timex` (*timex*)

Return a TimeContext object given a timex entry.

## Eidos Reader (`indra.sources.eidos.reader`)

**class** `indra.sources.eidos.reader.EidosReader`

Reader object keeping an instance of the Eidos reader as a singleton.

This allows the Eidos reader to need initialization when the first piece of text is read, the subsequent readings are done with the same instance of the reader and are therefore faster.

**eidos\_reader**

A Scala object, an instance of the Eidos reading system. It is instantiated only when first processing text.

**Type** `org.clulab.wm.eidos.EidosSystem`

**initialize\_reader** ()

Instantiate the Eidos reader attribute of this reader.

**process\_text** (*text, format='json'*)

Return a mentions JSON object given text.

**Parameters**

- **text** (*str*) – Text to be processed.
- **format** (*str*) – The format of the output to produce, one of “json” or “json\_ld”. Default: “json”

**Returns** `json_dict` – A JSON object of mentions extracted from text.

**Return type** `dict`

## Eidos CLI (`indra.sources.eidos.cli`)

This is a Python based command line interface to Eidos to complement the Python-Java bridge based interface. EIDOSPAC (in the INDRA config.ini or as an environmental variable) needs to be pointing to a fat JAR of the Eidos system.

`indra.sources.eidos.cli.extract_and_process(path_in, path_out)`

Run Eidos on a set of text files and process output with INDRA.

The output is produced in the specified output folder but the output files aren't processed by this function.

### Parameters

- **path\_in** (*str*) – Path to an input folder with some text files
- **path\_out** (*str*) – Path to an output folder in which Eidos places the output JSON-LD files

**Returns** `stmts` – A list of INDRA Statements

**Return type** `list[indra.statements.Statements]`

`indra.sources.eidos.cli.extract_from_directory(path_in, path_out)`

Run Eidos on a set of text files in a folder.

The output is produced in the specified output folder but the output files aren't processed by this function.

### Parameters

- **path\_in** (*str*) – Path to an input folder with some text files
- **path\_out** (*str*) – Path to an output folder in which Eidos places the output JSON-LD files

`indra.sources.eidos.cli.run_eidos(endpoint, *args)`

Run a given endpoint of Eidos through the command line.

### Parameters

- **endpoint** (*str*) – The class within the Eidos package to run, for instance 'apps.ExtractFromDirectory' will run 'org.clulab.wm.eidos.apps.ExtractFromDirectory'
- **\*args** – Any further arguments to be passed as inputs to the class being run.

## Eidos Webserver (`indra.sources.eidos.server`)

This is a Python-based web server that can be run to read with Eidos. To run the server, do

```
python -m indra.sources.eidos.server
```

and then submit POST requests to the `localhost:5000/process_text` endpoint with JSON content as `{'text': 'text to read'}`. The response will be the Eidos JSON-LD output.

## CWMS (`indra.sources.cwms`)

CWMS is a variant of the TRIPS system. It is a general purpose natural language understanding system with applications in world modeling. For more information, see: <http://trips.ihmc.us/parser/cgi/cwmsreader>

**CWMS API (`indra.sources.cwms.api`)**

`indra.sources.cwms.api.process_ekb(ekb_str)`

Processes an EKB string produced by CWMS.

**Parameters** `ekb_str` (*str*) – EKB string to process

**Returns** `cp` – A CWMSProcessor, which contains a list of INDRA statements in its statements attribute.

**Return type** `indra.sources.cwms.CWMSProcessor`

`indra.sources.cwms.api.process_ekb_file(fname)`

Processes an EKB file produced by CWMS.

**Parameters** `fname` (*str*) – Path to the EKB file to process.

**Returns** `cp` – A CWMSProcessor, which contains a list of INDRA statements in its statements attribute.

**Return type** `indra.sources.cwms.CWMSProcessor`

`indra.sources.cwms.api.process_rdf_file(text, rdf_filename)`

Process CWMS's RDF output for the given statement and returns a processor populated with INDRA statements.

**Parameters**

- `text` (*str*) – Sentence to process
- `rdf_filename` (*str*) – The RDF filename to process

**Returns** `cp` – A CWMSProcessor instance, which contains a list of INDRA Statements as its statements attribute.

**Return type** `indra.sources.cwms.CWMSRDFProcessor`

`indra.sources.cwms.api.process_text(text, save_xml='cwms_output.xml')`

Processes text using the CWMS web service.

**Parameters** `text` (*str*) – Text to process

**Returns** `cp` – A CWMSProcessor, which contains a list of INDRA statements in its statements attribute.

**Return type** `indra.sources.cwms.CWMSProcessor`

**CWMS EKB Processor (`indra.sources.cwms.processor`)**

**exception** `indra.sources.cwms.processor.CWMSError`

**class** `indra.sources.cwms.processor.CWMSProcessor(xml_string)`

The CWMSProcessor currently extracts causal relationships between terms (nouns) in EKB. In the future, this processor can be extended to extract other types of relations, or to extract relations involving events.

For more details on the TRIPS EKB XML format, see <http://trips.ihmc.us/parser/cgi/drum>

**Parameters** `xml_string` (*str*) – A TRIPS extraction knowledge base (EKB) in XML format as a string.

**tree**

An ElementTree object representation of the TRIPS EKB XML.

**Type** `xml.etree.ElementTree.Element`

**doc\_id**  
Document ID  
**Type** str

**statements**  
A list of INDRA Statements that were extracted from the EKB.  
**Type** list[indra.statements.Statement]

**sentences**  
The list of all sentences in the EKB with their IDs  
**Type** dict[str: str]

**paragraphs**  
The list of all paragraphs in the EKB with their IDs  
**Type** dict[str: str]

**par\_to\_sec**  
A map from paragraph IDs to their associated section types  
**Type** dict[str: str]

**event\_from\_event** (*event\_term*)  
Return an Event from an EVENT element in the EKB.

**extract\_causal\_relations** ()  
Extract Influence Statements from the EKB.

**extract\_events** ()  
Extract standalone Events from the EKB.

**influence\_from\_event** (*event*)  
Return an Influence from an EVENT element in the EKB.

**influence\_from\_relation** (*relation*)  
Return an Influence from a CC element in the EKB.

**migration\_from\_event** (*event\_term*)  
Return a Migration event from an EVENT element in the EKB.

### CWMS RDF Processor (`indra.sources.cwms.rdf_processor`)

**class** `indra.sources.cwms.rdf_processor.CWMSRDFProcessor` (*text*, *rdf\_filename*)

This processor extracts INDRA statements from CWMS RDF output.

#### Parameters

- **text** (*str*) – The source sentence as text
- **rdf\_filename** (*str*) – A string containing the RDF output returned by CWMS for that sentence

#### statements

A list of INDRA statements that were extracted by the processor.

**Type** list[indra.statements.Statement]

#### extract\_statement\_from\_query\_result

 (*res*)

Adds a statement based on one element of a rdflib SPARQL query.

**Parameters** **res** (*rdflib.query.ResultRow*) – Element of rdflib SPARQL query result

**extract\_statements ()**

Extracts INDRA statements from the RDF graph via SPARQL queries.

**Sofia (indra.sources.sofia)**

Sofia is a general purpose natural language processing system developed at UPitt and CMU by N. Miskov et al.

**Sofia API (indra.sources.sofia.api)**

`indra.sources.sofia.api.process_json (json_obj)`

Return processor by processing a JSON object returned by Sofia.

**Parameters** `json_obj (json)` – A JSON object containing extractions from Sofia.

**Returns** `sp` – A SofiaProcessor object which has a list of extracted INDRA Statements as its statements attribute.

**Return type** `indra.sources.sofia.processor.SofiaProcessor`

`indra.sources.sofia.api.process_json_file (fname)`

Return processor by processing a JSON file produced by Sofia.

**Parameters** `fname (str)` – The name of the JSON file to process

**Returns** A SofiaProcessor object which has a list of extracted INDRA Statements as its statements attribute.

**Return type** `indra.sources.sofia.processor.SofiaProcessor`

`indra.sources.sofia.api.process_table (fname)`

Return processor by processing a given sheet of a spreadsheet file.

**Parameters** `fname (str)` – The name of the Excel file (typically .xlsx extension) to process

**Returns** `sp` – A SofiaProcessor object which has a list of extracted INDRA Statements as its statements attribute.

**Return type** `indra.sources.sofia.processor.SofiaProcessor`

`indra.sources.sofia.api.process_text (text, out_file='sofia_output.json', auth=None)`

Return processor by processing text given as a string.

**Parameters**

- **text (str)** – A string containing the text to be processed with Sofia.
- **out\_file (Optional[str])** – The path to a file to save the reader's output into. Default: `sofia_output.json`
- **auth (Optional[list])** – A username/password pair for the Sofia web service. If not given, the `SOFIA_USERNAME` and `SOFIA_PASSWORD` values are loaded from either the INDRA config or the environment.

**Returns** `sp` – A SofiaProcessor object which has a list of extracted INDRA Statements as its statements attribute. If the API did not process the text, `None` is returned.

**Return type** `indra.sources.sofia.processor.SofiaProcessor`



**Sofia Processor** (`indra.sources.sofia.processor`)**Hume** (`indra.sources.hume`)

Hume is a general purpose reading system developed by BBN.

Currently, INDRA can process JSON-LD files produced by Hume. When available, the API will be extended with access to the reader as a service.

**Hume API** (`indra.sources.hume.api`)

`indra.sources.hume.api.process_jsonld_file` (*fname*)

Process a JSON-LD file in the new format to extract Statements.

**Parameters** *fname* (*str*) – The path to the JSON-LD file to be processed.

**Returns** A HumeProcessor instance, which contains a list of INDRA Statements as its statements attribute.

**Return type** `indra.sources.hume.HumeProcessor`

`indra.sources.hume.api.process_jsonld` (*jsonld*)

Process a JSON-LD string in the new format to extract Statements.

**Parameters** *jsonld* (*dict*) – The JSON-LD object to be processed.

**Returns** A HumeProcessor instance, which contains a list of INDRA Statements as its statements attribute.

**Return type** `indra.sources.hume.HumeProcessor`

**Hume Processor** (`indra.sources.hume.processor`)

**class** `indra.sources.hume.processor.HumeJsonLdProcessor` (*json\_dict*)

This processor extracts INDRA Statements from Hume JSON-LD output.

**Parameters** *json\_dict* (*dict*) – A JSON dictionary containing the Hume extractions in JSON-LD format.

**tree**

The objectpath Tree object representing the extractions.

**Type** `objectpath.Tree`

**statements**

A list of INDRA Statements that were extracted by the processor.

**Type** `list[indra.statements.Statement]`

## 4.2.3 Standard Molecular Pathway Databases

**BEL** (`indra.sources.bel`)**BEL API** (`indra.sources.bel.api`)

`indra.sources.bel.api.process_belrdf` (*rdf\_str*, *print\_output=True*)

Return a BelRdfProcessor for a BEL/RDF string.

**Parameters** `rdf_str` (*str*) – A BEL/RDF string to be processed. This will usually come from reading a .rdf file.

**Returns** `bp` – A BelRdfProcessor object which contains INDRA Statements in `bp.statements`.

**Return type** *BelRdfProcessor*

## Notes

This function calls all the specific `get_type_of_mechanism()` functions of the newly constructed `BelRdfProcessor` to extract INDRA Statements.

`indra.sources.bel.api.process_belscript` (*file\_name*, *\*\*kwargs*)

Return a `PybelProcessor` by processing a BEL script file.

Key word arguments are passed directly to `pybel.from_path`, for further information, see [pybel.readthedocs.io/en/latest/io.html#pybel.from\\_path](http://pybel.readthedocs.io/en/latest/io.html#pybel.from_path) Some keyword arguments we use here differ from the defaults of PyBEL, namely we set `citation_clearing` to `False` and `no_identifier_validation` to `True`.

**Parameters** `file_name` (*str*) – The path to a BEL script file.

**Returns** `bp` – A `PybelProcessor` object which contains INDRA Statements in `bp.statements`.

**Return type** *PybelProcessor*

`indra.sources.bel.api.process_cbn_jgif_file` (*file\_name*)

Return a `PybelProcessor` by processing a CBN JGIF JSON file.

**Parameters** `file_name` (*str*) – The path to a CBN JGIF JSON file.

**Returns** `bp` – A `PybelProcessor` object which contains INDRA Statements in `bp.statements`.

**Return type** *PybelProcessor*

`indra.sources.bel.api.process_json_file` (*file\_name*)

Return a `PybelProcessor` by processing a Node-Link JSON file.

For more information on this format, see: <http://pybel.readthedocs.io/en/latest/io.html#node-link-json>

**Parameters** `file_name` (*str*) – The path to a Node-Link JSON file.

**Returns** `bp` – A `PybelProcessor` object which contains INDRA Statements in `bp.statements`.

**Return type** *PybelProcessor*

`indra.sources.bel.api.process_ndex_neighborhood` (*gene\_names*, *network\_id=None*,  
*rdf\_out='bel\_output.rdf'*,  
*print\_output=True*)

Return a `BelRdfProcessor` for an NDEX network neighborhood.

### Parameters

- **gene\_names** (*list*) – A list of HGNC gene symbols to search the neighborhood of. Example: ['BRAF', 'MAP2K1']
- **network\_id** (*Optional[str]*) – The UUID of the network in NDEX. By default, the BEL Large Corpus network is used.
- **rdf\_out** (*Optional[str]*) – Name of the output file to save the RDF returned by the web service. This is useful for debugging purposes or to repeat the same query on an offline RDF file later. Default: `bel_output.rdf`

**Returns** `bp` – A `BelRdfProcessor` object which contains INDRA Statements in `bp.statements`.

**Return type** *BelRdfProcessor*

## Notes

This function calls `process_belrdf` to the returned RDF string from the webservice.

`indra.sources.bel.api.process_pybel_graph`

Return a `PybelProcessor` by processing a PyBEL graph.

**Parameters** `graph` (`pybel.struct.BELGraph`) – A PyBEL graph to process

**Returns** `bp` – A `PybelProcessor` object which contains INDRA Statements in `bp.statements`.

**Return type** `PybelProcessor`

`indra.sources.bel.api.process_pybel_neighborhood` (`gene_names`, `network_file=None`,  
`network_type='belscript'`,  
`**kwargs`)

Return `PybelProcessor` around neighborhood of given genes in a network.

This function processes the given network file and filters the returned Statements to ones that contain genes in the given list.

### Parameters

- **network\_file** (`Optional[str]`) – Path to the network file to process. If not given, by default, the BEL Large Corpus is used.
- **network\_type** (`Optional[str]`) – This function allows processing both BEL Script files and JSON files. This argument controls which type is assumed to be processed, and the value can be either 'belscript' or 'json'. Default: `bel_script`

**Returns** `bp` – A `PybelProcessor` object which contains INDRA Statements in `bp.statements`.

**Return type** `PybelProcessor`

## BEL RDF Processor (`indra.sources.bel.rdf_processor`)

**class** `indra.sources.bel.rdf_processor.BelRdfProcessor` (`g`)

The `BelRdfProcessor` extracts INDRA Statements from a BEL RDF model.

**Parameters** `g` (`rdflib.Graph`) – An RDF graph object containing the BEL model.

**g**

An RDF graph object containing the BEL model.

**Type** `rdflib.Graph`

**statements**

A list of extracted INDRA Statements representing direct mechanisms. This list should be used for assembly in INDRA.

**Type** `list[indra.statements.Statement]`

**indirect\_stmts**

A list of extracted INDRA Statements representing indirect mechanisms. This list should be used for assembly or model checking in INDRA.

**Type** `list[indra.statements.Statement]`

**converted\_direct\_stmts**

A list of all direct BEL statements, as strings, that were converted into INDRA Statements.

**Type** `list[str]`

**converted\_indirect\_stmts**

A list of all indirect BEL statements, as strings, that were converted into INDRA Statements.

**Type** list[str]

**degenerate\_stmts**

A list of degenerate BEL statements, as strings, in the BEL model.

**Type** list[str]

**all\_direct\_stmts**

A list of all BEL statements representing direct interactions, as strings, in the BEL model.

**Type** list[str]

**all\_indirect\_stmts**

A list of all BEL statements that represent indirect interactions, as strings, in the BEL model.

**Type** list[str]

**get\_activating\_mods ()**

Extract INDRA ActiveForm Statements with a single mod from BEL.

The SPARQL pattern used for extraction from BEL looks for a ModifiedProteinAbundance as subject and an Activity of a ProteinAbundance as object.

**Examples**

```
proteinAbundance(HGNC:INSR,proteinModification(P,Y))    directlyIncreases    kinaseActiv-
ity(proteinAbundance(HGNC:INSR))
```

**get\_activating\_subs ()**

Extract INDRA ActiveForm Statements based on a mutation from BEL.

The SPARQL pattern used to extract ActiveForm due to mutations look for a ProteinAbundance as a subject which has a child encoding the amino acid substitution. The object of the statement is an ActivityType of the same ProteinAbundance, which is either increased or decreased.

**Examples**

```
proteinAbundance(HGNC:NRAS,substitution(Q,61,K))    directlyIncreases    gtpBoundActiv-
ity(proteinAbundance(HGNC:NRAS))
```

```
proteinAbundance(HGNC:TP53,substitution(F,134,I))    directlyDecreases    transcriptionalActiv-
ity(proteinAbundance(HGNC:TP53))
```

**get\_activation ()**

Extract INDRA Inhibition/Activation Statements from BEL.

The SPARQL query used to extract Activation Statements looks for patterns in which the subject is an ActivityType (of a ProteinAbundance) or an Abundance (of a small molecule). The object has to be the ActivityType (typically of a ProteinAbundance) which is either increased or decreased.

**Examples**

```
abundance(CHEBI:gefitinib) directlyDecreases kinaseActivity(proteinAbundance(HGNC:EGFR))
```

```
kinaseActivity(proteinAbundance(HGNC:MAP3K5))    directlyIncreases    kinaseActiv-
ity(proteinAbundance(HGNC:MAP2K7))
```

This pattern covers the extraction of Gap/Gef and GtpActivation Statements, which are recognized by the object activity or the subject activity, respectively, being *gtpbound*.

### Examples

```
catalyticActivity(proteinAbundance(HGNC:RASA1))      directlyDecreases      gtpBoundActiv-
ity(proteinAbundance(PFH:"RAS Family"))
catalyticActivity(proteinAbundance(HGNC:SOS1))      directlyIncreases      gtpBoundActiv-
ity(proteinAbundance(HGNC:HRAS))
gtpBoundActivity(proteinAbundance(HGNC:HRAS))      directlyIncreases      catalyticActiv-
ity(proteinAbundance(HGNC:TIAM1))
```

#### **get\_all\_direct\_statements ()**

Get all directlyIncreases/Decreases BEL statements.

This method stores the results of the query in `self.all_direct_stmts` as a list of strings. The SPARQL query used to find direct BEL statements searches for all statements whose predicate is either `DirectlyIncreases` or `DirectlyDecreases`.

#### **get\_all\_indirect\_statements ()**

Get all indirect increases/decreases BEL statements.

This method stores the results of the query in `self.all_indirect_stmts` as a list of strings. The SPARQL query used to find indirect BEL statements searches for all statements whose predicate is either `Increases` or `Decreases`.

#### **get\_complexes ()**

Extract INDRA Complex Statements from BEL.

The SPARQL query used to extract Complexes looks for `ComplexAbundance` terms and their constituents. This pattern is distinct from other patterns in this processor in that it queries for terms, not full statements.

### Examples

```
complexAbundance(proteinAbundance(HGNC:PPARG), proteinAbundance(HGNC:RXRA)) decreases
biologicalProcess(MESHPP:"Insulin Resistance")
```

#### **get\_conversions ()**

Extract Conversion INDRA Statements from BEL.

The SPARQL query used to extract Conversions searches for a subject (controller) which is an `Abundance-Activity` which `directlyIncreases` a Reaction with a given list of Reactants and Products.

### Examples

```
catalyticActivity(proteinAbundance(HGNC:HMOX1))      directlyIncreases      reac-
tion(reactants(abundance(CHEBI:heme)),      products(abundance(SCHEM:Biliverdine),      abun-
dance(CHEBI:"carbon monoxide")))          dence(CHEBI:"carbon monoxide"))
```

#### **get\_degenerate\_statements ()**

Get all degenerate BEL statements.

Stores the results of the query in `self.degenerate_stmts`.

**get\_modifications ()**

Extract INDRA Modification Statements from BEL.

Two SPARQL patterns are used for extracting Modifications from BEL:

- q\_phospho1 assumes that the subject is an AbundanceActivity, which increases/decreases a Modified-ProteinAbundance.

Examples:

```
kinaseActivity(proteinAbundance(HGNC:IKBKE)) directlyIncreases proteinAbun-
dance(HGNC:IRF3,proteinModification(P,S,385))
```

```
phosphataseActivity(proteinAbundance(HGNC:DUSP4)) directlyDecreases proteinAbun-
dance(HGNC:MAPK1,proteinModification(P,T,185))
```

- q\_phospho2 assumes that the subject is a ProteinAbundance which increases/decreases a Modified-ProteinAbundance.

Examples:

```
proteinAbundance(HGNC:NGF) increases proteinAbun-
dance(HGNC:NFKBIA,proteinModification(P,Y,42))
```

```
proteinAbundance(HGNC:FGF1) decreases proteinAbun-
dance(HGNC:RB1,proteinModification(P))
```

**get\_transcription ()**

Extract Increase/DecreaseAmount INDRA Statements from BEL.

Three distinct SPARQL patterns are used to extract amount regulations from BEL.

- q\_tscript1 searches for a subject which is a Transcription ActivityType of a ProteinAbundance and an object which is an RNAAbundance that is either increased or decreased.

Examples:

```
transcriptionalActivity(proteinAbundance(HGNC:FOXP2)) directlyIncreases rnaAbun-
dance(HGNC:SYK)
```

```
transcriptionalActivity(proteinAbundance(HGNC:FOXP2)) directlyDecreases rnaAbun-
dance(HGNC:CALCRL)
```

- q\_tscript2 searches for a subject which is a ProteinAbundance and an object which is an RNAAbundance. Note that this pattern typically exists in an indirect form (i.e. increases/decreases).

Example:

```
proteinAbundance(HGNC:MTF1) directlyIncreases rnaAbundance(HGNC:LCN1)
```

- q\_tscript3 searches for a subject which is a ModifiedProteinAbundance, with an object which is an RNAAbundance. In the BEL large corpus, this pattern is found for subjects which are protein families or mouse/rat proteins, and the predicate in an indirect increase.

Example:

```
proteinAbundance(PFR:"Akt Family",proteinModification(P)) increases rnaAbun-
dance(RGD:Cald1)
```

**print\_statement\_coverage ()**

Display how many of the direct statements have been converted.

Also prints how many are considered 'degenerate' and not converted.

**print\_statements ()**

Print all extracted INDRA Statements.

`indra.sources.bel.rdf_processor.namespace_from_uri(uri)`  
 Return the entity namespace from the URI. Examples: [http://www.openbel.org/bel/p\\_HGNC\\_RAF1](http://www.openbel.org/bel/p_HGNC_RAF1) -> HGNC  
[http://www.openbel.org/bel/p\\_RGD\\_Raf1](http://www.openbel.org/bel/p_RGD_Raf1) -> RGD [http://www.openbel.org/bel/p\\_PFH\\_MEK1/2\\_Family](http://www.openbel.org/bel/p_PFH_MEK1/2_Family) -> PFH

`indra.sources.bel.rdf_processor.term_from_uri(uri)`  
 Removes prepended URI information from terms.

### PyBEL Processor (`indra.sources.bel.processor`)

**class** `indra.sources.bel.processor.PybelProcessor` (*graph*)

Extract INDRA Statements from a PyBEL Graph.

Currently does not handle non-causal relationships (positiveCorrelation, (negativeCorrelation, hasVariant, etc.)

**Parameters** `graph` (*pybel.BELGraph*) – PyBEL graph containing the BEL content.

**statements**

A list of extracted INDRA Statements representing BEL Statements.

**Type** `list[indra.statements.Statement]`

### Biopax (`indra.sources.biopax`)

#### Biopax API (`indra.sources.biopax.api`)

`indra.sources.biopax.api.process_pc_neighborhood` (*gene\_names*, *neighbor\_limit=1*,  
*database\_filter=None*)

Returns a BiopaxProcessor for a PathwayCommons neighborhood query.

The neighborhood query finds the neighborhood around a set of source genes.

<http://www.pathwaycommons.org/pc2/#graph>

[http://www.pathwaycommons.org/pc2/#graph\\_kind](http://www.pathwaycommons.org/pc2/#graph_kind)

**Parameters**

- **gene\_names** (*list*) – A list of HGNC gene symbols to search the neighborhood of. Examples: ['BRAF'], ['BRAF', 'MAP2K1']
- **neighbor\_limit** (*Optional[int]*) – The number of steps to limit the size of the neighborhood around the gene names being queried. Default: 1
- **database\_filter** (*Optional[list]*) – A list of database identifiers to which the query is restricted. Examples: ['reactome'], ['biogrid', 'pid', 'psp'] If not given, all databases are used in the query. For a full list of databases see <http://www.pathwaycommons.org/pc2/datasources>

**Returns** `bp` – A BiopaxProcessor containing the obtained BioPAX model in `bp.model`.

**Return type** *BiopaxProcessor*

`indra.sources.biopax.api.process_pc_pathsbetween` (*gene\_names*, *neighbor\_limit=1*,  
*database\_filter=None*,  
*block\_size=None*)

Returns a BiopaxProcessor for a PathwayCommons paths-between query.

The paths-between query finds the paths between a set of genes. Here source gene names are given in a single list and all directions of paths between these genes are considered.

<http://www.pathwaycommons.org/pc2/#graph>

[http://www.pathwaycommons.org/pc2/#graph\\_kind](http://www.pathwaycommons.org/pc2/#graph_kind)

#### Parameters

- **gene\_names** (*list*) – A list of HGNC gene symbols to search for paths between. Examples: ['BRAF', 'MAP2K1']
- **neighbor\_limit** (*Optional[int]*) – The number of steps to limit the length of the paths between the gene names being queried. Default: 1
- **database\_filter** (*Optional[list]*) – A list of database identifiers to which the query is restricted. Examples: ['reactome'], ['biogrid', 'pid', 'psp'] If not given, all databases are used in the query. For a full list of databases see <http://www.pathwaycommons.org/pc2/datasources>
- **block\_size** (*Optional[int]*) – Large paths-between queries (above ~60 genes) can error on the server side. In this case, the query can be replaced by a series of smaller paths-between and paths-from-to queries each of which contains block\_size genes.

**Returns** **bp** – A BiopaxProcessor containing the obtained BioPAX model in bp.model.

**Return type** *BiopaxProcessor*

```
indra.sources.biopax.api.process_pc_pathsfromto(source_genes, target_genes, neighbor_limit=1, database_filter=None)
```

Returns a BiopaxProcessor for a PathwayCommons paths-from-to query.

The paths-from-to query finds the paths from a set of source genes to a set of target genes.

<http://www.pathwaycommons.org/pc2/#graph>

[http://www.pathwaycommons.org/pc2/#graph\\_kind](http://www.pathwaycommons.org/pc2/#graph_kind)

#### Parameters

- **source\_genes** (*list*) – A list of HGNC gene symbols that are the sources of paths being searched for. Examples: ['BRAF', 'RAF1', 'ARAF']
- **target\_genes** (*list*) – A list of HGNC gene symbols that are the targets of paths being searched for. Examples: ['MAP2K1', 'MAP2K2']
- **neighbor\_limit** (*Optional[int]*) – The number of steps to limit the length of the paths between the source genes and target genes being queried. Default: 1
- **database\_filter** (*Optional[list]*) – A list of database identifiers to which the query is restricted. Examples: ['reactome'], ['biogrid', 'pid', 'psp'] If not given, all databases are used in the query. For a full list of databases see <http://www.pathwaycommons.org/pc2/datasources>

**Returns** **bp** – A BiopaxProcessor containing the obtained BioPAX model in bp.model.

**Return type** *BiopaxProcessor*

```
indra.sources.biopax.api.process_owl(owl_filename)
```

Returns a BiopaxProcessor for a BioPAX OWL file.

**Parameters** **owl\_filename** (*str*) – The name of the OWL file to process.

**Returns** **bp** – A BiopaxProcessor containing the obtained BioPAX model in bp.model.

**Return type** *BiopaxProcessor*

```
indra.sources.biopax.api.process_owl_str(owl_str)
```

Returns a BiopaxProcessor for a BioPAX OWL file.



**Parameters** `owl_str` (*str*) – The string content of an OWL file to process.

**Returns** `bp` – A BiopaxProcessor containing the obtained BioPAX model in `bp.model`.

**Return type** *BiopaxProcessor*

`indra.sources.biopax.api.process_model(model)`

Returns a BiopaxProcessor for a BioPAX model object.

**Parameters** `model` (*org.biopax.paxtools.model.Model*) – A BioPAX model object.

**Returns** `bp` – A BiopaxProcessor containing the obtained BioPAX model in `bp.model`.

**Return type** *BiopaxProcessor*

## Biopax Processor (`indra.sources.biopax.processor`)

**class** `indra.sources.biopax.processor.BiopaxProcessor(model)`

The BiopaxProcessor extracts INDRA Statements from a BioPAX model.

The BiopaxProcessor uses pattern searches in a BioPAX OWL model to extract mechanisms from which it constructs INDRA Statements.

**Parameters** `model` (*org.biopax.paxtools.model.Model*) – A BioPAX model object  
(java object)

**model**

A BioPAX model object (java object) which is queried using Paxtools to extract INDRA Statements

**Type** `org.biopax.paxtools.model.Model`

**statements**

A list of INDRA Statements that were extracted from the model.

**Type** `list[indra.statements.Statement]`

**eliminate\_exact\_duplicates()**

Eliminate Statements that were extracted multiple times.

Due to the way the patterns are implemented, they can sometimes yield the same Statement information multiple times, in which case, we end up with redundant Statements that aren't from independent underlying entries. To avoid this, here, we filter out such duplicates.

**get\_activity\_modification()**

Extract INDRA ActiveForm statements from the BioPAX model.

This method extracts ActiveForm Statements that are due to protein modifications. This method reuses the structure of BioPAX Pattern's `org.biopax.paxtools.pattern.PatternBox.constrolsStateChange` pattern with additional constraints to specify the gain or loss of a modification occurring (phosphorylation, deubiquitination, etc.) and the gain or loss of activity due to the modification state change.

**get\_complexes()**

Extract INDRA Complex Statements from the BioPAX model.

This method searches for `org.biopax.paxtools.model.level3.Complex` objects which represent molecular complexes. It doesn't reuse BioPAX Pattern's `org.biopax.paxtools.pattern.PatternBox.inComplexWith` query since that retrieves pairs of complex members rather than the full complex.

**get\_conversions()**

Extract Conversion INDRA Statements from the BioPAX model.

This method uses a custom BioPAX Pattern (one that is not implemented `PatternBox`) to query for BiochemicalReactions whose left and right hand sides are collections of `SmallMolecules`. This pattern thereby

extracts metabolic conversions as well as signaling processes via small molecules (e.g. lipid phosphorylation or cleavage).

**get\_gap()**

Extract Gap INDRA Statements from the BioPAX model.

This method uses a custom BioPAX Pattern (one that is not implemented PatternBox) to query for controlled BiochemicalReactions in which the same protein is in complex with GTP on the left hand side and in complex with GDP on the right hand side. This implies that the controller is a GAP for the GDP/GTP-bound protein.

**get\_gef()**

Extract Gef INDRA Statements from the BioPAX model.

This method uses a custom BioPAX Pattern (one that is not implemented PatternBox) to query for controlled BiochemicalReactions in which the same protein is in complex with GDP on the left hand side and in complex with GTP on the right hand side. This implies that the controller is a GEF for the GDP/GTP-bound protein.

**get\_modifications()**

Extract INDRA Modification Statements from the BioPAX model.

To extract Modifications, this method reuses the structure of BioPAX Pattern's `org.biopax.paxtools.pattern.PatternBox.constrolsStateChange` pattern with additional constraints to specify the type of state change occurring (phosphorylation, deubiquitination, etc.).

**get\_regulate\_activities()**

Get Activation/Inhibition INDRA Statements from the BioPAX model.

This method extracts Activation/Inhibition Statements and reuses the structure of BioPAX Pattern's `org.biopax.paxtools.pattern.PatternBox.constrolsStateChange` pattern with additional constraints to specify the gain or loss of activity state but assuring that the activity change is not due to a modification state change (which are extracted by `get_modifications` and `get_activity_modification`).

**get\_regulate\_amounts()**

Extract INDRA RegulateAmount Statements from the BioPAX model.

This method extracts IncreaseAmount/DecreaseAmount Statements from the BioPAX model. It fully reuses BioPAX Pattern's `org.biopax.paxtools.pattern.PatternBox.controlsExpressionWithTemplateReac` pattern to find TemplateReactions which control the expression of a protein.

**print\_statements()**

Print all INDRA Statements collected by the processors.

**save\_model** (*file\_name=None*)

Save the BioPAX model object in an OWL file.

**Parameters** `file_name` (*Optional[str]*) – The name of the OWL file to save the model in.

`indra.sources.biopax.processor.to_impl(x)`

Return an Impl instance casted version the given biopax object.

**Pathway Commons Client** (`indra.sources.biopax.pathway_commons_client`)

`indra.sources.biopax.pathway_commons_client.graph_query` (*kind, source, target=None, neighbor\_limit=1, database\_filter=None*)

Perform a graph query on PathwayCommons.

For more information on these queries, see <http://www.pathwaycommons.org/pc2/#graph>

**Parameters**

- **kind** (*str*) – The kind of graph query to perform. Currently 3 options are implemented, ‘neighborhood’, ‘pathsbetween’ and ‘pathsfromto’.
- **source** (*list[str]*) – A list of gene names which are the source set for the graph query.
- **target** (*Optional[list[str]]*) – A list of gene names which are the target set for the graph query. Only needed for ‘pathsfromto’ queries.
- **neighbor\_limit** (*Optional[int]*) – This limits the length of the longest path considered in the graph query. Default: 1

**Returns** **model** – A BioPAX model (java object).

**Return type** `org.biopax.paxtools.model.Model`

`indra.sources.biopax.pathway_commons_client.model_to_owl(model, fname)`  
Save a BioPAX model object as an OWL file.

**Parameters**

- **model** (*org.biopax.paxtools.model.Model*) – A BioPAX model object (java object).
- **fname** (*str*) – The name of the OWL file to save the model in.

`indra.sources.biopax.pathway_commons_client.owl_str_to_model(owl_str)`  
Return a BioPAX model object from an OWL string.

**Parameters** **owl\_str** (*str*) – The model as an OWL string.

**Returns** **biopax\_model** – A BioPAX model object (java object).

**Return type** `org.biopax.paxtools.model.Model`

`indra.sources.biopax.pathway_commons_client.owl_to_model(fname)`  
Return a BioPAX model object from an OWL file.

**Parameters** **fname** (*str*) – The name of the OWL file containing the model.

**Returns** **biopax\_model** – A BioPAX model object (java object).

**Return type** `org.biopax.paxtools.model.Model`

**SIGNOR** (`indra.sources.signor`)**SIGNOR API** (`indra.sources.signor.api`)

`indra.sources.signor.api.process_from_file(signor_data_file, signor_complexes_file=None)`

Process Signor interaction data from CSV files.

**Parameters**

- **signor\_data\_file** (*str*) – Path to the Signor interaction data file in CSV format.
- **signor\_complexes\_file** (*str*) – Path to the Signor complexes data in CSV format. If unspecified, Signor complexes will not be expanded to their constituents.

**Returns** SignorProcessor containing Statements extracted from the Signor data.

**Return type** `indra.sources.signor.SignorProcessor`

**SIGNOR Processor (`indra.sources.signor.processor`)**

An input processor for the SIGNOR database: a database of causal relationships between biological entities.

See publication:

Perfetto et al., “SIGNOR: a database of causal relationships between biological entities,” *Nucleic Acids Research*, Volume 44, Issue D1, 4 January 2016, Pages D548-D554. <https://doi.org/10.1093/nar/gkv1048>

**class** `indra.sources.signor.processor.SignorProcessor` (*data*, *complex\_map=None*)  
Processor for Signor dataset, available at <http://signor.uniroma2.it>.

**Parameters**

- **data** (*iterator*) – Iterator over rows of a SIGNOR CSV file.
- **complex\_map** (*dict*) – A dict containing SIGNOR complexes, keyed by their IDs.

**statements**

A list of INDRA Statements extracted from the SIGNOR table.

**Type** `list[indra.statements.Statements]`

**no\_mech\_rows**

List of rows where no mechanism statements were generated.

**Type** `list` of `SignorRow` namedtuples

**no\_mech\_ctr**

Counter listing the frequency of different MECHANISM types in the list of no-mechanism rows.

**Type** `collections.Counter`

**BioGrid (`indra.sources.biogrid`)**

**class** `indra.sources.biogrid.BiogridProcessor` (*biogrid\_file=None*, *physical\_only=True*)  
Extracts INDRA Complex statements from Biogrid interaction data.

**Parameters**

- **biogrid\_file** (*str*) – The file containing the Biogrid data in .tab2 format. If not provided, the BioGrid data is downloaded from the BioGrid website.
- **physical\_only** (*boolean*) – If True, only physical interactions are included (e.g., genetic interactions are excluded). If False, all interactions are included).

**statements**

Extracted INDRA Complex statements.

**Type** `list[indra.statements.Statements]`

**physical\_only**

Indicates whether only physical interactions were included during statement processing.

**Type** `boolean`

**Human Protein Reference Database (`indra.sources.hprd`)**

This module implements getting content from the Human Protein Reference Database (HPRD), a curated protein data resource, as INDRA Statements. In particular, the module supports extracting post-translational modifications, protein complexes, and (binary) protein-protein interactions from HPRD.

More information about HPRD can be obtained at <http://www.hprd.org> and in these publications:

- Peri, S. et al. (2003). Development of Human Protein Reference Database as an initial platform for approaching systems biology in humans. *Genome Research*. 13, 2363-2371.
- Prasad, T. S. K. et al. (2009). Human Protein Reference Database - 2009 Update. *Nucleic Acids Research*. 37, D767-72.

Data from the final release of HPRD (version 9) can be obtained at the following URLs:

- [http://www.hprd.org/RELEASE9/HPRD\\_FLAT\\_FILES\\_041310.tar.gz](http://www.hprd.org/RELEASE9/HPRD_FLAT_FILES_041310.tar.gz) (text files)
- [http://www.hprd.org/RELEASE9/HPRD\\_XML\\_041310.tar.gz](http://www.hprd.org/RELEASE9/HPRD_XML_041310.tar.gz) (XML)

This module is designed to process the text files obtained from the first link listed above.

### HPRD API (`indra.sources.hprd.api`)

```
indra.sources.hprd.api.process_flat_files(id_mappings_file, complexes_file=None,
                                          ptm_file=None, ppi_file=None, seq_file=None,
                                          motif_window=7)
```

Get INDRA Statements from HPRD data.

Of the arguments, *id\_mappings\_file* is required, and at least one of *complexes\_file*, *ptm\_file*, and *ppi\_file* must also be given. If *ptm\_file* is given, *seq\_file* must also be given.

Note that many proteins (> 1,600) in the HPRD content are associated with outdated RefSeq IDs that cannot be mapped to Uniprot IDs. For these, the Uniprot ID obtained from the HGNC ID (itself obtained from the Entrez ID) is used. Because the sequence referenced by the Uniprot ID obtained this way may be different from the (outdated) RefSeq sequence included with the HPRD content, it is possible that this will lead to invalid site positions with respect to the Uniprot IDs.

To allow these site positions to be mapped during assembly, the Modification statements produced by the Hprd-Processor include an additional key in the *annotations* field of their Evidence object. The annotations field is called 'site\_motif' and it maps to a dictionary with three elements: 'motif', 'respos', and 'off\_by\_one'. 'motif' gives the peptide sequence obtained from the RefSeq sequence included with HPRD. 'respos' indicates the position in the peptide sequence containing the residue. Note that these positions are ONE-INDEXXED (not zero-indexed). Finally, the 'off-by-one' field contains a boolean value indicating whether the correct position was inferred as being an off-by-one (methionine cleavage) error. If True, it means that the given residue could not be found in the HPRD RefSeq sequence at the given position, but a matching residue was found at position+1, suggesting a sequence numbering based on the methionine-cleaved sequence. The peptide included in the 'site\_motif' dictionary is based on this updated position.

#### Parameters

- **id\_mappings\_file** (*str*) – Path to HPRD\_ID\_MAPPINGS.txt file.
- **complexes\_file** (*Optional[str]*) – Path to PROTEIN\_COMPLEXES.txt file.
- **ptm\_file** (*Optional[str]*) – Path to POST\_TRANSLATIONAL\_MODIFICATIONS.txt file.
- **ppi\_file** (*Optional[str]*) – Path to BINARY\_PROTEIN\_PROTEIN\_INTERACTIONS.txt file.
- **seq\_file** (*Optional[str]*) – Path to PROTEIN\_SEQUENCES.txt file.
- **motif\_window** (*int*) – Number of flanking amino acids to include on each side of the PTM target residue in the 'site\_motif' annotations field of the Evidence for Modification Statements. Default is 7.

**Returns** An HprdProcessor object which contains a list of extracted INDRA Statements in its statements attribute.

**Return type** *HprdProcessor*

### HPRD Processor (`indra.sources.hprd.processor`)

```
class indra.sources.hprd.processor.HprdProcessor (id_df, cplx_df=None, ptm_df=None,
                                                    ppi_df=None, seq_dict=None, motif_window=7)
```

Get INDRA Statements from HPRD data.

See documentation for `indra.sources.hprd.api.process_flat_files`.

#### Parameters

- **id\_df** (*pandas.DataFrame*) – DataFrame loaded from the HPRD\_ID\_MAPPINGS.txt file.
- **cplx\_df** (*pandas.DataFrame*) – DataFrame loaded from the PROTEIN\_COMPLEXES.txt file.
- **ptm\_df** (*pandas.DataFrame*) – DataFrame loaded from the POST\_TRANSLATIONAL\_MODIFICATIONS.txt file.
- **ppi\_df** (*pandas.DataFrame*) – DataFrame loaded from the BINARY\_PROTEIN\_PROTEIN\_INTERACTIONS.txt file.
- **seq\_dict** (*dict*) – Dictionary mapping RefSeq IDs to protein sequences, loaded from the PROTEIN\_SEQUENCES.txt file.
- **motif\_window** (*int*) – Number of flanking amino acids to include on each side of the PTM target residue in the ‘site\_motif’ annotations field of the Evidence for Modification Statements. Default is 7.

#### statements

INDRA Statements (Modifications and Complexes) produced from the HPRD content.

**Type** list of INDRA Statements

#### id\_df

DataFrame loaded from HPRD\_ID\_MAPPINGS.txt file.

**Type** pandas.DataFrame

#### seq\_dict

Dictionary mapping RefSeq IDs to protein sequences, loaded from the PROTEIN\_SEQUENCES.txt file.

#### no\_hgnc\_for\_egid

Counter listing Entrez gene IDs reference in the HPRD content that could not be mapped to a current HGNC ID, along with their frequency.

**Type** collections.Counter

#### no\_up\_for\_hgnc

Counter with tuples of form (entrez\_id, hgnc\_symbol, hgnc\_id) where the HGNC ID could not be mapped to a Uniprot ID, along with their frequency.

**Type** collections.Counter

#### no\_up\_for\_refseq

Counter of RefSeq protein IDs that could not be mapped to any Uniprot ID, along with frequency.

**Type** collections.Counter

**many\_ups\_for\_refseq**

Counter of RefSeq protein IDs that yielded more than one matching Uniprot ID. Note that in these cases, the Uniprot ID obtained from HGNC is used.

**Type** collections.Counter

**invalid\_site\_pos**

List of tuples of form (refseq\_id, residue, position) indicating sites of post translational modifications where the protein sequences provided by HPRD did not contain the given residue at the given position.

**Type** list of tuples

**off\_by\_one**

The subset of sites contained in *invalid\_site\_pos* where the given residue can be found at position+1 in the HPRD protein sequence, suggesting an off-by-one error due to numbering based on the protein with initial methionine cleaved. Note that no mapping is performed by the processor.

**Type** list of tuples

**motif\_window**

Number of flanking amino acids to include on each side of the PTM target residue in the 'site\_motif' annotations field of the Evidence for Modification Statements. Default is 7.

**Type** int

**get\_complexes** (*cplx\_df*)

Generate Complex Statements from the HPRD protein complexes data.

**Parameters** **cplx\_df** (*pandas.DataFrame*) – DataFrame loaded from the PROTEIN\_COMPLEXES.txt file.

**get\_ppis** (*ppi\_df*)

Generate Complex Statements from the HPRD PPI data.

**Parameters** **ppi\_df** (*pandas.DataFrame*) – DataFrame loaded from the BINARY\_PROTEIN\_PROTEIN\_INTERACTIONS.txt file.

**get\_ptms** (*ptm\_df*)

Generate Modification statements from the HPRD PTM data.

**Parameters** **ptm\_df** (*pandas.DataFrame*) – DataFrame loaded from the POST\_TRANSLATIONAL\_MODIFICATIONS.txt file.

## TRRUST Database (`indra.sources.trrust`)

This module provides an interface to the TRRUST knowledge base and extracts TF-target relationships as INDRA Statements.

TRRUST is available at <https://www.grnpedia.org/trrust/>, see also <https://www.ncbi.nlm.nih.gov/pubmed/29087512>.

## TRRUST API (`indra.sources.trrust.api`)

`indra.sources.trrust.api.process_from_web()`

Return a TrrustProcessor based on the online interaction table.

**Returns** A TrrustProcessor object that has a list of INDRA Statements in its statements attribute.

**Return type** *TrrustProcessor*

### TRRUST Processor (`indra.sources.trrust.processor`)

**class** `indra.sources.trrust.processor.TrrustProcessor` (*df*)

Processor to extract INDRA Statements from Trrust data frame.

**df**

The Trrust table to process.

**Type** `pandas.DataFrame`

**statements**

The list of INDRA Statements extracted from the table.

**Type** `list[indra.statements.Statement]`

**extract\_statements** ()

Process the table to extract Statements.

`indra.sources.trrust.processor.get_grounded_agent` (*gene\_name*)

Return a grounded Agent based on an HGNC symbol.

`indra.sources.trrust.processor.make_stmt` (*stmt\_cls, tf\_agent, target\_agent, pmid*)

Return a Statement based on its type, agents, and PMID.

## 4.2.4 Custom Knowledge Bases

### Target Affinity Spectrum (`indra.sources.tas`)

This module provides an API and processor to the Target Affinity Spectrum data set compiled by N. Moret in the Laboratory of Systems Pharmacology at HMS. This data set is based on experiments as opposed to the manually curated drug-target relationships provided in the LINCS small molecule dataset.

Moret, N., et al. (2018). Cheminformatics tools for analyzing and designing optimized small molecule libraries. *BioRxiv*, (617), 358978. <https://doi.org/10.1101/358978>

### TAS API (`indra.sources.tas.api`)

`indra.sources.tas.api.process_csv` (*affinity\_class\_limit=2*)

Return a `TasProcessor` for the contents of the csv contained in data.

**Interactions are classified into the following classes based on affinity:**

1 –  $K_d < 100\text{nM}$

2 –  $100\text{nM} < K_d < 1\mu\text{M}$

3 –  $1\mu\text{M} < K_d < 10\mu\text{M}$

10 –  $K_d > 10\mu\text{M}$

By default, only classes 1 and 2 are extracted but the `affinity_class_limit` parameter can be used to change the upper limit of extracted classes.

**Parameters** `affinity_class_limit` (*Optional[int]*) – Defines the highest class of binding affinity that is included in the extractions. Default: 2

**Returns** A `TasProcessor` object which has a list of INDRA Statements extracted from the CSV file representing drug-target inhibitions in its `statements` attribute.

**Return type** `TasProcessor`



**TAS Processor** (`indra.sources.tas.processor`)

**class** `indra.sources.tas.processor.TasProcessor` (*data*, *affinity\_class\_limit*)  
 A processor for the Target Affinity Spectrum data table.

**LINCS Drug Targets** (`indra.sources.lincs_drug`)

This module provides an API and processor for the HMS LINCS small molecule target relationship database. This is a manually curated set of relationships with the “nominal” target of each drug determined by a human expert. Note that the determination of the “nominal” target is not always backed up by experimentally measured affinities. The underlying data is available here: <http://lincs.hms.harvard.edu/db/datasets/20000/results>

**LINCS Drug Targets API** (`indra.sources.lincs_drug.api`)

`indra.sources.lincs_drug.api.process_from_web` ()  
 Return a processor for the LINCS drug target data.

**Returns** A `LincsProcessor` object which contains extracted INDRA Statements in its `statements` attribute.

**Return type** *LincsProcessor*

**LINCS Drug Targets Processor** (`indra.sources.lincs_drug.processor`)

**class** `indra.sources.lincs_drug.processor.LincsProcessor` (*lincs\_data*)  
 Processor for the HMS LINCS drug target dataset.

**Parameters** `lincs_data` (*list[dict]*) – A list of dicts with keys set by the header of the csv, and values from the data in the csv.

**statements**

A list of indra statements extracted from the CSV file.

**Type** `list[indra.statements.Statement]`

**NDEX CX API** (`indra.sources.ndex_cx.api`)

`indra.sources.ndex_cx.api.process_cx` (*cx\_json*, *summary=None*, *require\_grounding=True*)  
 Process a CX JSON object into Statements.

**Parameters**

- **cx\_json** (*list*) – CX JSON object.
- **summary** (*Optional[dict]*) – The network summary object which can be obtained via `get_network_summary` through the web service. This contains metadata such as the owner and the creation time of the network.
- **require\_grounding** (*bool*) – Whether network nodes lacking grounding information should be included among the extracted Statements (default is `True`).

**Returns** Processor containing Statements.

**Return type** *NdexCxProcessor*

`indra.sources.ndex_cx.api.process_cx_file` (*file\_name*, *require\_grounding=True*)  
 Process a CX JSON file into Statements.

**Parameters**

- **file\_name** (*str*) – Path to file containing CX JSON.
- **require\_grounding** (*bool*) – Whether network nodes lacking grounding information should be included among the extracted Statements (default is True).

**Returns** Processor containing Statements.

**Return type** *NdexCxProcessor*

`indra.sources.ndex_cx.api.process_ndex_network` (*network\_id*, *username=None*, *password=None*, *require\_grounding=True*)  
 Process an NDEx network into Statements.

**Parameters**

- **network\_id** (*str*) – NDEx network ID.
- **username** (*str*) – NDEx username.
- **password** (*str*) – NDEx password.
- **require\_grounding** (*bool*) – Whether network nodes lacking grounding information should be included among the extracted Statements (default is True).

**Returns** Processor containing Statements. Returns None if there if the HTTP status code indicates an unsuccessful request.

**Return type** *NdexCxProcessor*

**NDEx CX Processor (`indra.sources.ndex_cx.processor`)**

`class indra.sources.ndex_cx.processor.NdexCxProcessor` (*cx*, *summary=None*, *require\_grounding=True*)

The NdexCxProcessor extracts INDRA Statements from Cytoscape CX JSON.

**Parameters**

- **cx** (*list of dicts*) – JSON content containing the Cytoscape network in CX format.
- **summary** (*Optional[dict]*) – The network summary object which can be obtained via `get_network_summary` through the web service. This contains metadata such as the owner and the creation time of the network.

**statements**

A list of extracted INDRA Statements. Not all edges in the network may be converted into Statements.

**Type** list

**get\_agents** ()

Get list of grounded nodes in the network as Agents.

**Returns** Only nodes containing sufficient information to be grounded will be contained in this list.

**Return type** list of Agents

**get\_node\_names** ()

Get list of all nodes in the network by name.

**get\_pmids ()**  
Get list of all PMIDs associated with edges in the network.

**get\_statements ()**  
Convert network edges into Statements.

**Returns** Converted INDRA Statements.

**Return type** list of Statements

### INDRA Database REST Client (`indra.sources.indra_db_rest`)

The INDRA database client allows querying a web service that serves content from a database of INDRA Statements collected and pre-assembled from various sources.

Access to the webservice requires a URL (`INDRA_DB_REST_URL`) and possibly an API key (`INDRA_DB_REST_API_KEY`), both of which may be placed in your config file or as environment variables. If you do not have these but would like to access the database REST API, you may contact the developers to request a URL and API key.

### INDRA Database REST API (`indra.sources.indra_db_rest.api`)

```
indra.sources.indra_db_rest.api.get_statements (subject=None,          object=None,
                                               agents=None,          stmt_type=None,
                                               use_exact_type=False,  persist=True,
                                               timeout=None, simple_response=False,
                                               ev_limit=10, best_first=True, tries=2,
                                               max_stmts=None)
```

Get a processor for the INDRA DB web API matching given agents and type.

There are two types of responses available. You can just get a list of INDRA Statements, or you can get an `IndraDBRestProcessor` object, which allow Statements to be loaded in a background thread, providing a sample of the best\* content available promptly in the `sample_statements` attribute, and populates the `statements` attribute when the paged load is complete.

The latter should be used in all new code, and where convenient the prior should be converted to use the processor, as this option may be removed in the future.

- In the sense of having the most supporting evidence.

#### Parameters

- **subject/object** (*str*) – Optionally specify the subject and/or object of the statements in you wish to get from the database. By default, the namespace is assumed to be HGNC gene names, however you may specify another namespace by including `@<namespace>` at the end of the name string. For example, if you want to specify an agent by chebi, you could use `CHEBI:6801@CHEBI`, or if you wanted to use the HGNC id, you could use `6871@HGNC`.
- **agents** (*list[str]*) – A list of agents, specified in the same manner as subject and object, but without specifying their grammatical position.
- **stmt\_type** (*str*) – Specify the types of interactions you are interested in, as indicated by the sub-classes of INDRA's Statements. This argument is *not* case sensitive. If the statement class given has sub-classes (e.g. `RegulateAmount` has `IncreaseAmount` and `DecreaseAmount`), then both the class itself, and its subclasses, will be queried, by default. If you do not want this behavior, set `use_exact_type=True`. Note that if `max_stmts` is set, it

is possible only the exact statement type will be returned, as this is the first searched. The processor then cycles through the types, getting a page of results for each type and adding it to the quota, until the max number of statements is reached.

- **use\_exact\_type** (*bool*) – If *stmt\_type* is given, and you only want to search for that specific statement type, set this to True. Default is False.
- **persist** (*bool*) – Default is True. When False, if a query comes back limited (not all results returned), just give up and pass along what was returned. Otherwise, make further queries to get the rest of the data (which may take some time).
- **timeout** (*positive int or None*) – If an int, block until the work is done and statements are retrieved, or until the timeout has expired, in which case the results so far will be returned in the response object, and further results will be added in a separate thread as they become available. If *simple\_response* is True, all statements available will be returned. Otherwise (if None), block indefinitely until all statements are retrieved. Default is None.
- **simple\_response** (*bool*) – If True, a simple list of statements is returned (thus block should also be True). If block is False, only the original sample will be returned (as though persist was False), until the statements are done loading, in which case the rest should appear in the list. This behavior is not encouraged. Default is False (which breaks backwards compatibility with usage of INDRA versions from before 1/22/2019). WE ENCOURAGE ALL NEW USE-CASES TO USE THE PROCESSOR, AS THIS FEATURE MAY BE REMOVED AT A LATER DATE.
- **ev\_limit** (*int or None*) – Limit the amount of evidence returned per Statement. Default is 10.
- **best\_first** (*bool*) – If True, the preassembled statements will be sorted by the amount of evidence they have, and those with the most evidence will be prioritized. When using *max\_stmts*, this means you will get the “best” statements. If False, statements will be queried in arbitrary order.
- **tries** (*int > 0*) – Set the number of times to try the query. The database often caches results, so if a query times out the first time, trying again after a timeout will often succeed fast enough to avoid a timeout. This can also help gracefully handle an unreliable connection, if you’re willing to wait. Default is 2.
- **max\_stmts** (*int or None*) – Select the maximum number of statements to return. When set less than 1000 the effect is much the same as setting persist to false, and will guarantee a faster response. Default is None.

**Returns processor** – An instance of the `IndraDBRestProcessor`, which has an attribute *statements* which will be populated when the query/queries are done. This is the default behavior, and is encouraged in all future cases, however a simple list of statements may be returned using the *simple\_response* option described above.

**Return type** `IndraDBRestProcessor`

```
indra.sources.indra_db_rest.api.get_statements_for_paper(ids, ev_limit=10,
                                                         best_first=True, tries=2,
                                                         max_stmts=None)
```

Get the set of raw Statements extracted from a paper given by the id.

#### Parameters

- **ids** (*list[(`<id type>`, `<id value>`)]*) – A list of tuples with ids and their type. The type can be any one of ‘pmid’, ‘pmcid’, ‘doi’, ‘pii’, ‘manuscript id’, or ‘trid’, which is the primary key id of the text references in the database.

- **ev\_limit** (*int* or *None*) – Limit the amount of evidence returned per Statement. Default is 10.
- **best\_first** (*bool*) – If True, the preassembled statements will be sorted by the amount of evidence they have, and those with the most evidence will be prioritized. When using *max\_stmts*, this means you will get the “best” statements. If False, statements will be queried in arbitrary order.
- **tries** (*int* > 0) – Set the number of times to try the query. The database often caches results, so if a query times out the first time, trying again after a timeout will often succeed fast enough to avoid a timeout. This can also help gracefully handle an unreliable connection, if you’re willing to wait. Default is 2.
- **max\_stmts** (*int* or *None*) – Select a maximum number of statements to be returned. Default is None.

**Returns** *stmts* – A list of INDRA Statement instances.

**Return type** `list[indra.statements.Statement]`

```
indra.sources.indra_db_rest.api.get_statements_by_hash(hash_list, ev_limit=100,
                                                       best_first=True, tries=2)
```

Get fully formed statements from a list of hashes.

#### Parameters

- **hash\_list** (*list[int or str]*) – A list of statement hashes.
- **ev\_limit** (*int* or *None*) – Limit the amount of evidence returned per Statement. Default is 100.
- **best\_first** (*bool*) – If True, the preassembled statements will be sorted by the amount of evidence they have, and those with the most evidence will be prioritized. When using *max\_stmts*, this means you will get the “best” statements. If False, statements will be queried in arbitrary order.
- **tries** (*int* > 0) – Set the number of times to try the query. The database often caches results, so if a query times out the first time, trying again after a timeout will often succeed fast enough to avoid a timeout. This can also help gracefully handle an unreliable connection, if you’re willing to wait. Default is 2.

```
indra.sources.indra_db_rest.api.submit_curation(hash_val, tag, curator, text=None,
                                                source='indra_rest_client',
                                                ev_hash=None, is_test=False)
```

Submit a curation for the given statement at the relevant level.

#### Parameters

- **hash\_val** (*int*) – The hash corresponding to the statement.
- **tag** (*str*) – A very short phrase categorizing the error or type of curation, e.g. “grounding” for a grounding error, or “correct” if you are marking a statement as correct.
- **curator** (*str*) – The name or identifier for the curator.
- **text** (*str*) – A brief description of the problem.
- **source** (*str*) – The name of the access point through which the curation was performed. The default is ‘direct\_client’, meaning this function was used directly. Any higher-level application should identify itself here.
- **ev\_hash** (*int*) – A hash of the sentence and other evidence information. Elsewhere referred to as *source\_hash*.

- **is\_test** (*bool*) – Used in testing. If True, no curation will actually be added to the database.

### INDRA Database REST Processor (`indra.sources.indra_db_rest.processor`)

```
class indra.sources.indra_db_rest.processor.IndraDBRestProcessor (subject=None,  
object=None,  
agents=None,  
stmt_type=None,  
use_exact_type=False,  
persist=True,  
time-  
out=None,  
ev_limit=10,  
best_first=True,  
tries=2,  
max_stmts=None)
```

The packaging for query responses.

#### Parameters

- **subject/object** (*str*) – Optionally specify the subject and/or object of the statements in you wish to get from the database. By default, the namespace is assumed to be HGNC gene names, however you may specify another namespace by including `@<namespace>` at the end of the name string. For example, if you want to specify an agent by chebi, you could use `CHEBI:6801@CHEBI`, or if you wanted to use the HGNC id, you could use `6871@HGNC`.
- **agents** (*list[str]*) – A list of agents, specified in the same manner as subject and object, but without specifying their grammatical position.
- **stmt\_type** (*str*) – Specify the types of interactions you are interested in, as indicated by the sub-classes of INDRA's Statements. This argument is *not* case sensitive. If the statement class given has sub-classes (e.g. RegulateAmount has IncreaseAmount and DecreaseAmount), then both the class itself, and its subclasses, will be queried, by default. If you do not want this behavior, set `use_exact_type=True`. Note that if `max_stmts` is set, it is possible only the exact statement type will be returned, as this is the first searched. The processor then cycles through the types, getting a page of results for each type and adding it to the quota, until the max number of statements is reached.
- **use\_exact\_type** (*bool*) – If `stmt_type` is given, and you only want to search for that specific statement type, set this to True. Default is False.
- **persist** (*bool*) – Default is True. When False, if a query comes back limited (not all results returned), just give up and pass along what was returned. Otherwise, make further queries to get the rest of the data (which may take some time).
- **timeout** (*positive int or None*) – If an int, block until the work is done and statements are retrieved, or until the timeout has expired, in which case the results so far will be returned in the response object, and further results will be added in a separate thread as they become available. If `simple_response` is True, all statements available will be returned. Otherwise (if None), block indefinitely until all statements are retrieved. Default is None.
- **ev\_limit** (*int or None*) – Limit the amount of evidence returned per Statement. Default is 10.

- **best\_first** (*bool*) – If True, the preassembled statements will be sorted by the amount of evidence they have, and those with the most evidence will be prioritized. When using *max\_stmts*, this means you will get the “best” statements. If False, statements will be queried in arbitrary order.
- **tries** (*int > 0*) – Set the number of times to try the query. The database often caches results, so if a query times out the first time, trying again after a timeout will often succeed fast enough to avoid a timeout. This can also help gracefully handle an unreliable connection, if you’re willing to wait. Default is 2.
- **max\_stmts** (*int or None*) – Select the maximum number of statements to return. When set less than 1000 the effect is much the same as setting *persist* to false, and will guarantee a faster response. Default is None.

**statements**

A list of INDRA Statements that will be filled once all queries have been completed.

**Type** list[indra.statements.Statement]

**statements\_sample**

A list of the INDRA Statements received from the first query. In general these will be the “best” (currently this means they have the most evidence) Statements available.

**Type** list[indra.statements.Statement]

**get\_ev\_count** (*stmt*)

Get the total evidence count for a statement.

**get\_ev\_count\_by\_hash** (*stmt\_hash*)

Get the total evidence count for a statement hash.

**get\_ev\_counts** ()

Get a dictionary of evidence counts.

**get\_hash\_statements\_dict** ()

Return a dict of Statements keyed by hashes.

**get\_source\_count** (*stmt*)

Get the source counts for a given statement.

**get\_source\_count\_by\_hash** (*stmt\_hash*)

Get the source counts for a given statement.

**get\_source\_counts** ()

Get the source counts as a dict per statement hash.

**is\_working** ()

Check if the thread is running.

**merge\_results** (*other\_processor*)

Merge the results of this processor with those of another.

**wait\_until\_done** (*timeout=None*)

Wait for the background load to complete.

## 4.3 Database clients (`indra.databases`)

### 4.3.1 HGNC client (`indra.hgnc_client`)

`indra.databases.hgnc_client.get_current_hgnc_id(hgnc_name)`

Return the HGNC ID(s) corresponding to a current or outdated HGNC symbol.

**Parameters** `hgnc_name` (*str*) – The HGNC symbol to be converted, possibly an outdated symbol.

**Returns** If there is a single HGNC ID corresponding to the given current or outdated HGNC symbol, that ID is returned as a string. If the symbol is outdated and maps to multiple current IDs, a list of these IDs is returned. If the given name doesn't correspond to either a current or an outdated HGNC symbol, `None` is returned.

**Return type** `str` or list of `str` or `None`

`indra.databases.hgnc_client.get_entrez_id(hgnc_id)`

Return the Entrez ID corresponding to the given HGNC ID.

**Parameters** `hgnc_id` (*str*) – The HGNC ID to be converted. Note that the HGNC ID is a number that is passed as a string. It is not the same as the HGNC gene symbol.

**Returns** `entrez_id` – The Entrez ID corresponding to the given HGNC ID.

**Return type** `str`

`indra.databases.hgnc_client.get_hgnc_entry`

Return the HGNC entry for the given HGNC ID from the web service.

**Parameters** `hgnc_id` (*str*) – The HGNC ID to be converted.

**Returns** `xml_tree` – The XML ElementTree corresponding to the entry for the given HGNC ID.

**Return type** `ElementTree`

`indra.databases.hgnc_client.get_hgnc_from_entrez(entrez_id)`

Return the HGNC ID corresponding to the given Entrez ID.

**Parameters** `entrez_id` (*str*) – The EntrezC ID to be converted, a number passed as a string.

**Returns** `hgnc_id` – The HGNC ID corresponding to the given Entrez ID.

**Return type** `str`

`indra.databases.hgnc_client.get_hgnc_from_mouse(mgi_id)`

Return the HGNC ID corresponding to the given MGI mouse gene ID.

**Parameters** `mgi_id` (*str*) – The MGI ID to be converted. Example: "2444934"

**Returns** `hgnc_id` – The HGNC ID corresponding to the given MGI ID.

**Return type** `str`

`indra.databases.hgnc_client.get_hgnc_from_rat(rgd_id)`

Return the HGNC ID corresponding to the given RGD rat gene ID.

**Parameters** `rgd_id` (*str*) – The RGD ID to be converted. Example: "1564928"

**Returns** `hgnc_id` – The HGNC ID corresponding to the given RGD ID.

**Return type** `str`

`indra.databases.hgnc_client.get_hgnc_id(hgnc_name)`

Return the HGNC ID corresponding to the given HGNC symbol.



**Parameters** `hgnc_name` (*str*) – The HGNC symbol to be converted. Example: BRAF

**Returns** `hgnc_id` – The HGNC ID corresponding to the given HGNC symbol.

**Return type** `str`

`indra.databases.hgnc_client.get_hgnc_name(hgnc_id)`

Return the HGNC symbol corresponding to the given HGNC ID.

**Parameters** `hgnc_id` (*str*) – The HGNC ID to be converted.

**Returns** `hgnc_name` – The HGNC symbol corresponding to the given HGNC ID.

**Return type** `str`

`indra.databases.hgnc_client.get_mouse_id(hgnc_id)`

Return the MGI mouse ID corresponding to the given HGNC ID.

**Parameters** `hgnc_id` (*str*) – The HGNC ID to be converted. Example: “”

**Returns** `mgid_id` – The MGI ID corresponding to the given HGNC ID.

**Return type** `str`

`indra.databases.hgnc_client.get_rat_id(hgnc_id)`

Return the RGD rat ID corresponding to the given HGNC ID.

**Parameters** `hgnc_id` (*str*) – The HGNC ID to be converted. Example: “”

**Returns** `rgd_id` – The RGD ID corresponding to the given HGNC ID.

**Return type** `str`

`indra.databases.hgnc_client.get_uniprot_id(hgnc_id)`

Return the UniProt ID corresponding to the given HGNC ID.

**Parameters** `hgnc_id` (*str*) – The HGNC ID to be converted. Note that the HGNC ID is a number that is passed as a string. It is not the same as the HGNC gene symbol.

**Returns** `uniprot_id` – The UniProt ID corresponding to the given HGNC ID.

**Return type** `str`

`indra.databases.hgnc_client.is_kinase(gene_name)`

Return True if the given gene name is a kinase.

**Parameters** `gene_name` (*str*) – The HGNC gene symbol corresponding to the protein.

**Returns** True if the given gene name corresponds to a kinase, False otherwise.

**Return type** `bool`

`indra.databases.hgnc_client.is_phosphatase(gene_name)`

Return True if the given gene name is a phosphatase.

**Parameters** `gene_name` (*str*) – The HGNC gene symbol corresponding to the protein.

**Returns** True if the given gene name corresponds to a phosphatase, False otherwise.

**Return type** `bool`

`indra.databases.hgnc_client.is_transcription_factor(gene_name)`

Return True if the given gene name is a transcription factor.

**Parameters** `gene_name` (*str*) – The HGNC gene symbol corresponding to the protein.

**Returns** True if the given gene name corresponds to a transcription factor, False otherwise.

**Return type** `bool`

### 4.3.2 Uniprot client (`indra.databases.uniprot_client`)

### 4.3.3 ChEBI client (`indra.databases.chebi_client`)

`indra.databases.chebi_client.get_chebi_entry_from_web`

Return a ChEBI entry corresponding to a given ChEBI ID using a REST API.

**Parameters** `chebi_id` (*str*) – The ChEBI ID whose entry is to be returned.

**Returns** An ElementTree element representing the ChEBI entry.

**Return type** `xml.etree.ElementTree.Element`

`indra.databases.chebi_client.get_chebi_id_from_cas` (*cas\_id*)

Return a ChEBI ID corresponding to the given CAS ID.

**Parameters** `cas_id` (*str*) – The CAS ID to be converted.

**Returns** `chebi_id` – The ChEBI ID corresponding to the given CAS ID. If the lookup fails, None is returned.

**Return type** `str`

`indra.databases.chebi_client.get_chebi_id_from_hmdb` (*hmdb\_id*)

Return the ChEBI ID corresponding to an HMDB ID.

**Parameters** `hmdb_id` (*str*) – An HMDB ID.

**Returns** The ChEBI ID that the given HMDB ID maps to or None if no mapping was found.

**Return type** `str`

`indra.databases.chebi_client.get_chebi_id_from_name` (*chebi\_name*)

Return a ChEBI ID corresponding to the given ChEBI name.

**Parameters** `chebi_name` (*str*) – The ChEBI name whose ID is to be returned.

**Returns** `chebi_id` – The ID corresponding to the given ChEBI name. If the lookup fails, None is returned.

**Return type** `str`

`indra.databases.chebi_client.get_chebi_id_from_pubchem` (*pubchem\_id*)

Return the ChEBI ID corresponding to a given Pubchem ID.

**Parameters** `pubchem_id` (*str*) – Pubchem ID to be converted.

**Returns** `chebi_id` – ChEBI ID corresponding to the given Pubchem ID. If the lookup fails, None is returned.

**Return type** `str`

`indra.databases.chebi_client.get_chebi_name_from_id` (*chebi\_id*, *offline=False*)

Return a ChEBI name corresponding to the given ChEBI ID.

**Parameters**

- `chebi_id` (*str*) – The ChEBI ID whose name is to be returned.
- `offline` (*Optional[bool]*) – Choose whether to allow an online lookup if the local lookup fails. If True, the online lookup is not attempted. Default: False.

**Returns** `chebi_name` – The name corresponding to the given ChEBI ID. If the lookup fails, None is returned.

**Return type** `str`

`indra.databases.chebi_client.get_chebi_name_from_id_web(chebi_id)`

Return a ChEBI name corresponding to a given ChEBI ID using a REST API.

**Parameters** `chebi_id` (*str*) – The ChEBI ID whose name is to be returned.

**Returns** `chebi_name` – The name corresponding to the given ChEBI ID. If the lookup fails, None is returned.

**Return type** `str`

`indra.databases.chebi_client.get_chembl_id(chebi_id)`

Return a ChEMBL ID from a given ChEBI ID.

**Parameters** `chebi_id` (*str*) – ChEBI ID to be converted.

**Returns** `chembl_id` – ChEMBL ID corresponding to the given ChEBI ID. If the lookup fails, None is returned.

**Return type** `str`

`indra.databases.chebi_client.get_inchi_key(chebi_id)`

Return an InChIKey corresponding to a given ChEBI ID using a REST API.

**Parameters** `chebi_id` (*str*) – The ChEBI ID whose InChIKey is to be returned.

**Returns** The InChIKey corresponding to the given ChEBI ID. If the lookup fails, None is returned.

**Return type** `str`

`indra.databases.chebi_client.get_pubchem_id(chebi_id)`

Return the PubChem ID corresponding to a given ChEBI ID.

**Parameters** `chebi_id` (*str*) – ChEBI ID to be converted.

**Returns** `pubchem_id` – PubChem ID corresponding to the given ChEBI ID. If the lookup fails, None is returned.

**Return type** `str`

`indra.databases.chebi_client.get_specific_id(chebi_ids)`

Return the most specific ID in a list based on the hierarchy.

**Parameters** `chebi_ids` (*list of str*) – A list of ChEBI IDs some of which may be hierarchically related.

**Returns** The first ChEBI ID which is at the most specific level in the hierarchy with respect to the input list.

**Return type** `str`

#### 4.3.4 BioGRID client (`indra.databases.biogrid_client`)

`indra.databases.biogrid_client.get_publications(gene_names, save_json_name=None)`

Return evidence publications for interaction between the given genes.

**Parameters**

- **gene\_names** (*list[str]*) – A list of gene names (HGNC symbols) to query interactions between. Currently supports exactly two genes only.
- **save\_json\_name** (*Optional[str]*) – A file name to save the raw BioGRID web service output in. By default, the raw output is not saved.

**Returns** `publications` – A list of Publication objects that provide evidence for interactions between the given list of genes.

**Return type** list[Publication]

### 4.3.5 Cell type context client (`indra.databases.context_client`)

`indra.databases.context_client.get_mutations` (*gene\_names*, *cell\_types*)

Return protein amino acid changes in given genes and cell types.

**Parameters**

- **gene\_names** (*list*) – HGNC gene symbols for which mutations are queried.
- **cell\_types** (*list*) – List of cell type names in which mutations are queried. The cell type names follow the CCLE database conventions.

Example: LOXIMVI\_SKIN, BT20\_BREAST

**Returns** **res** – A dictionary keyed by cell line, which contains another dictionary that is keyed by gene name, with a list of amino acid substitutions as values.

**Return type** dict[dict[list]]

`indra.databases.context_client.get_protein_expression` (*gene\_names*, *cell\_types*)

Return the protein expression levels of genes in cell types.

**Parameters**

- **gene\_names** (*list*) – HGNC gene symbols for which expression levels are queried.
- **cell\_types** (*list*) – List of cell type names in which expression levels are queried. The cell type names follow the CCLE database conventions.

Example: LOXIMVI\_SKIN, BT20\_BREAST

**Returns** **res** – A dictionary keyed by cell line, which contains another dictionary that is keyed by gene name, with estimated protein amounts as values.

**Return type** dict[dict[float]]

### 4.3.6 Network relevance client (`indra.databases.relevance_client`)

`indra.databases.relevance_client.get_heat_kernel` (*network\_id*)

Return the identifier of a heat kernel calculated for a given network.

**Parameters** **network\_id** (*str*) – The UUID of the network in NDEx.

**Returns** **kernel\_id** – The identifier of the heat kernel calculated for the given network.

**Return type** str

`indra.databases.relevance_client.get_relevant_nodes` (*network\_id*, *query\_nodes*)

Return a set of network nodes relevant to a given query set.

A heat diffusion algorithm is used on a pre-computed heat kernel for the given network which starts from the given query nodes. The nodes in the network are ranked according to heat score which is a measure of relevance with respect to the query nodes.

**Parameters**

- **network\_id** (*str*) – The UUID of the network in NDEx.
- **query\_nodes** (*list[str]*) – A list of node names with respect to which relevance is queried.

**Returns** `ranked_entities` – A list containing pairs of node names and their relevance scores.

**Return type** `list[(str, float)]`

### 4.3.7 NDEx client (`indra.databases.ndex_client`)

`indra.databases.ndex_client.create_network(cx_str, ndex_cred=None, private=True)`  
Creates a new NDEx network of the assembled CX model.

To upload the assembled CX model to NDEx, you need to have a registered account on NDEx (<http://ndexbio.org/>) and have the `ndex` python package installed. The uploaded network is private by default.

**Parameters** `ndex_cred` (*dict*) – A dictionary with the following entries: ‘user’: NDEx user name ‘password’: NDEx password

**Returns** `network_id` – The UUID of the NDEx network that was created by uploading the assembled CX model.

**Return type** `str`

`indra.databases.ndex_client.get_default_ndex_cred(ndex_cred)`  
Gets the NDEx credentials from the dict, or tries the environment if None

`indra.databases.ndex_client.send_request(ndex_service_url, params, is_json=True, use_get=False)`

Send a request to the NDEx server.

#### Parameters

- `ndex_service_url` (*str*) – The URL of the service to use for the request.
- `params` (*dict*) – A dictionary of parameters to send with the request. Parameter keys differ based on the type of request.
- `is_json` (*bool*) – True if the response is in json format, otherwise it is assumed to be text. Default: False
- `use_get` (*bool*) – True if the request needs to use GET instead of POST.

**Returns** `res` – Depending on the type of service and the `is_json` parameter, this function either returns a text string or a json dict.

**Return type** `str`

`indra.databases.ndex_client.set_style(network_id, ndex_cred=None, template_id=None)`  
Set the style of the network to a given template network’s style

#### Parameters

- `network_id` (*str*) – The UUID of the NDEx network whose style is to be changed.
- `ndex_cred` (*dict*) – A dictionary of NDEx credentials.
- `template_id` (*Optional[str]*) – The UUID of the NDEx network whose style is used on the network specified in the first argument.

`indra.databases.ndex_client.update_network(cx_str, network_id, ndex_cred=None)`  
Update an existing CX network on NDEx with new CX content.

#### Parameters

- `cx_str` (*str*) – String containing the CX content.
- `network_id` (*str*) – UUID of the network on NDEx.

- **ndex\_cred** (*dict*) – A dictionary with the following entries: ‘user’: NDEx user name  
‘password’: NDEx password

### 4.3.8 cBio portal client (`indra.databases.cbio_client`)

`indra.databases.cbio_client.get_cancer_studies` (*study\_filter=None*)

Return a list of cancer study identifiers, optionally filtered.

There are typically multiple studies for a given type of cancer and a filter can be used to constrain the returned list.

**Parameters** **study\_filter** (*Optional[str]*) – A string used to filter the study IDs to return.

Example: “paad”

**Returns** **study\_ids** – A list of study IDs. For instance “paad” as a filter would result in a list of study IDs with paad in their name like “paad\_icgc”, “paad\_tcga”, etc.

**Return type** list[str]

`indra.databases.cbio_client.get_cancer_types` (*cancer\_filter=None*)

Return a list of cancer types, optionally filtered.

**Parameters** **cancer\_filter** (*Optional[str]*) – A string used to filter cancer types. Its value is the name or part of the name of a type of cancer. Example: “melanoma”, “pancreatic”, “non-small cell lung”

**Returns** **type\_ids** – A list of cancer types matching the filter. Example: for `cancer_filter=“pancreatic”`, the result includes “panet” (neuro-endocrine) and “paad” (adenocarcinoma)

**Return type** list[str]

`indra.databases.cbio_client.get_case_lists` (*study\_id*)

Return a list of the case set ids for a particular study.

TAKE NOTE the “case\_list\_id” are the same thing as “case\_set\_id” Within the data, this string is referred to as a “case\_list\_id”. Within API calls it is referred to as a ‘case\_set\_id’. The documentation does not make this explicitly clear.

**Parameters** **study\_id** (*str*) – The ID of the cBio study. Example: ‘cellline\_ccle\_broad’ or ‘paad\_icgc’

**Returns** **case\_set\_ids** – A dict keyed to cases containing a dict keyed to genes containing int

**Return type** dict[dict[int]]

`indra.databases.cbio_client.get_ccle_cna` (*gene\_list, cell\_lines*)

Return a dict of CNAs in given genes and cell lines from CCLE.

CNA values correspond to the following alterations

-2 = homozygous deletion

-1 = hemizygous deletion

0 = neutral / no change

1 = gain

2 = high level amplification

**Parameters**

- **gene\_list** (*list[str]*) – A list of HGNC gene symbols to get mutations in

- **cell\_lines** (*list[str]*) – A list of CCLE cell line names to get mutations for.

**Returns** **profile\_data** – A dict keyed to cases containing a dict keyed to genes containing int

**Return type** dict[dict[int]]

```
indra.databases.cbio_client.get_ccle_lines_for_mutation(gene,
                                                    amino_acid_change)
```

Return cell lines with a given point mutation in a given gene.

Checks which cell lines in CCLE have a particular point mutation in a given gene and return their names in a list.

#### Parameters

- **gene** (*str*) – The HGNC symbol of the mutated gene in whose product the amino acid change occurs. Example: “BRAF”
- **amino\_acid\_change** (*str*) – The amino acid change of interest. Example: “V600E”

**Returns** **cell\_lines** – A list of CCLE cell lines in which the given mutation occurs.

**Return type** list

```
indra.databases.cbio_client.get_ccle_mrna(gene_list, cell_lines)
```

Return a dict of mRNA amounts in given genes and cell lines from CCLE.

#### Parameters

- **gene\_list** (*list[str]*) – A list of HGNC gene symbols to get mRNA amounts for.
- **cell\_lines** (*list[str]*) – A list of CCLE cell line names to get mRNA amounts for.

**Returns** **mrna\_amounts** – A dict keyed to cell lines containing a dict keyed to genes containing float

**Return type** dict[dict[float]]

```
indra.databases.cbio_client.get_ccle_mutations(gene_list, cell_lines, mutation_type=None)
```

Return a dict of mutations in given genes and cell lines from CCLE.

This is a specialized call to `get_mutations` tailored to CCLE cell lines.

#### Parameters

- **gene\_list** (*list[str]*) – A list of HGNC gene symbols to get mutations in
- **cell\_lines** (*list[str]*) – A list of CCLE cell line names to get mutations for.
- **mutation\_type** (*Optional[str]*) – The type of mutation to filter to. `mutation_type` can be one of: `missense`, `nonsense`, `frame_shift_ins`, `frame_shift_del`, `splice_site`

#### Returns

**mutations** – The result from cBioPortal as a dict in the format {`cell_line` : {`gene` : [`mutation1`, `mutation2`, ...]}}

Example: {‘LOXIMVI\_SKIN’: {‘BRAF’: [‘V600E’, ‘I208V’]}, ‘SKMEL30\_SKIN’: {‘BRAF’: [‘D287H’, ‘E275K’]}}

**Return type** dict

```
indra.databases.cbio_client.get_genetic_profiles(study_id, profile_filter=None)
```

Return all the genetic profiles (data sets) for a given study.

Genetic profiles are different types of data for a given study. For instance the study ‘cellline\_ccle\_broad’ has profiles such as ‘cellline\_ccle\_broad\_mutations’ for mutations, ‘cellline\_ccle\_broad\_CNA’ for copy number alterations, etc.

**Parameters**

- **study\_id** (*str*) – The ID of the cBio study. Example: ‘paad\_icgc’
- **profile\_filter** (*Optional[str]*) – A string used to filter the profiles to return. Will be one of: - MUTATION - MUTATION\_EXTENDED - COPY\_NUMBER\_ALTERATION - MRNA\_EXPRESSION - METHYLATION The genetic profiles can include “mutation”, “CNA”, “rppa”, “methylation”, etc.

**Returns** **genetic\_profiles** – A list of genetic profiles available for the given study.

**Return type** list[str]

```
indra.databases.cbio_client.get_mutations(study_id, gene_list, mutation_type=None,
                                         case_id=None)
```

Return mutations as a list of genes and list of amino acid changes.

**Parameters**

- **study\_id** (*str*) – The ID of the cBio study. Example: ‘cellline\_ccle\_broad’ or ‘paad\_icgc’
- **gene\_list** (*list[str]*) – A list of genes with their HGNC symbols. Example: [‘BRAF’, ‘KRAS’]
- **mutation\_type** (*Optional[str]*) – The type of mutation to filter to. *mutation\_type* can be one of: missense, nonsense, frame\_shift\_ins, frame\_shift\_del, splice\_site
- **case\_id** (*Optional[str]*) – The case ID within the study to filter to.

**Returns** **mutations** – A tuple of two lists, the first one containing a list of genes, and the second one a list of amino acid changes in those genes.

**Return type** tuple[list]

```
indra.databases.cbio_client.get_num_sequenced(study_id)
```

Return number of sequenced tumors for given study.

This is useful for calculating mutation statistics in terms of the prevalence of certain mutations within a type of cancer.

**Parameters** **study\_id** (*str*) – The ID of the cBio study. Example: ‘paad\_icgc’

**Returns** **num\_case** – The number of sequenced tumors in the given study

**Return type** int

```
indra.databases.cbio_client.get_profile_data(study_id, gene_list, profile_filter,
                                             case_set_filter=None)
```

Return dict of cases and genes and their respective values.

**Parameters**

- **study\_id** (*str*) – The ID of the cBio study. Example: ‘cellline\_ccle\_broad’ or ‘paad\_icgc’
- **gene\_list** (*list[str]*) – A list of genes with their HGNC symbols. Example: [‘BRAF’, ‘KRAS’]
- **profile\_filter** (*str*) – A string used to filter the profiles to return. Will be one of: - MUTATION - MUTATION\_EXTENDED - COPY\_NUMBER\_ALTERATION - MRNA\_EXPRESSION - METHYLATION



- **case\_set\_filter** (*Optional[str]*) – A string that specifies which case\_set\_id to use, based on a complete or partial match. If not provided, will look for study\_id + ‘\_all’

**Returns profile\_data** – A dict keyed to cases containing a dict keyed to genes containing int

**Return type** dict[dict[int]]

`indra.databases.cbio_client.send_request`

Return a data frame from a web service request to cBio portal.

Sends a web service request to the cBio portal with arguments given in the dictionary data and returns a Pandas data frame on success.

More information about the service here: [http://www.cbioportal.org/web\\_api.jsp](http://www.cbioportal.org/web_api.jsp)

**Parameters kwargs** (*dict*) – A dict of parameters for the query. Entries map directly to web service calls with the exception of the optional ‘skiprows’ entry, whose value is used as the number of rows to skip when reading the result data frame.

**Returns df** – Response from cBioPortal as a Pandas DataFrame.

**Return type** pandas.DataFrame

### 4.3.9 ChEMBL client (`indra.databases.chembl_client`)

`indra.databases.chembl_client.activities_by_target` (*activities*)

Get back lists of activities in a dict keyed by ChEMBL target id

**Parameters activities** (*list*) – response from a query returning activities for a drug

**Returns targ\_act\_dict** – dictionary keyed to ChEMBL target ids with lists of activity ids

**Return type** dict

`indra.databases.chembl_client.get_chembl_id` (*nlm\_mesh*)

Get ChEMBL ID from NLM MESH

**Parameters nlm\_mesh** (*str*) –

**Returns chembl\_id**

**Return type** str

`indra.databases.chembl_client.get_drug_inhibition_stmts` (*drug*)

Query ChEMBL for kinetics data given drug as Agent get back statements

**Parameters drug** (*Agent*) – Agent representing drug with MESH or CHEBI grounding

**Returns stmts** – INDRA statements generated by querying ChEMBL for all kinetics data of a drug interacting with protein targets

**Return type** list of INDRA statements

`indra.databases.chembl_client.get_evidence` (*assay*)

Given an activity, return an INDRA Evidence object.

**Parameters assay** (*dict*) – an activity from the activities list returned by a query to the API

**Returns ev** – an Evidence object containing the kinetics of the

**Return type** Evidence

`indra.databases.chembl_client.get_kinetics` (*assay*)

Given an activity, return its kinetics values.

**Parameters assay** (*dict*) – an activity from the activities list returned by a query to the API

**Returns** **kin** – dictionary of values with units keyed to value types ‘IC50’, ‘EC50’, ‘INH’, ‘Potency’, ‘Kd’

**Return type** dict

`indra.databases.chembl_client.get_mesh_id(nlm_mesh)`

Get MESH ID from NLM MESH

**Parameters** **nlm\_mesh** (*str*) –

**Returns** **mesh\_id**

**Return type** str

`indra.databases.chembl_client.get_pcid(mesh_id)`

Get PC ID from MESH ID

**Parameters** **mesh** (*str*) –

**Returns** **pcid**

**Return type** str

`indra.databases.chembl_client.get_pmid(doc_id)`

Get PMID from document\_chembl\_id

**Parameters** **doc\_id** (*str*) –

**Returns** **pmid**

**Return type** str

`indra.databases.chembl_client.get_protein_targets_only(target_chembl_ids)`

Given list of ChEMBL target ids, return dict of SINGLE PROTEIN targets

**Parameters** **target\_chembl\_ids** (*list*) – list of chembl\_ids as strings

**Returns** **protein\_targets** – dictionary keyed to ChEMBL target ids with lists of activity ids

**Return type** dict

`indra.databases.chembl_client.get_target_chemblid(target_upid)`

Get ChEMBL ID from UniProt upid

**Parameters** **target\_upid** (*str*) –

**Returns** **target\_chembl\_id**

**Return type** str

`indra.databases.chembl_client.query_target(target_chembl_id)`

Query ChEMBL API target by id

**Parameters** **target\_chembl\_id** (*str*) –

**Returns** **target** – dict parsed from json that is unique for the target

**Return type** dict

`indra.databases.chembl_client.send_query(query_dict)`

Query ChEMBL API

**Parameters** **query\_dict** (*dict*) – ‘query’: string of the endpoint to query ‘params’: dict of params for the query

**Returns** **js** – dict parsed from json that is unique to the submitted query

**Return type** dict

### 4.3.10 LINC client (`indra.databases.lincs_client`)

`indra.databases.lincs_client.get_drug_target_data()`

Load the csv into a list of dicts containing the LINC drug target data.

**Returns data** – A list of dicts, each keyed based on the header of the csv, with values as the corresponding column values.

**Return type** list[dict]

**class** `indra.databases.lincs_client.LincsClient`

Client for querying LINC small molecules and proteins.

`get_protein_refs(hms_lincs_id)`

Get the refs for a protein from the LINC protein metadata.

**Parameters** `hms_lincs_id` (*str*) – The HMS LINC ID for the protein

**Returns** A dictionary of protein references.

**Return type** dict

`get_small_molecule_name(hms_lincs_id)`

Get the name of a small molecule from the LINC sm metadata.

**Parameters** `hms_lincs_id` (*str*) – The HMS LINC ID of the small molecule.

**Returns** The name of the small molecule.

**Return type** str

`get_small_molecule_refs(hms_lincs_id)`

Get the id refs of a small molecule from the LINC sm metadata.

**Parameters** `hms_lincs_id` (*str*) – The HMS LINC ID of the small molecule.

**Returns** A dictionary of references.

**Return type** dict

`indra.databases.lincs_client.load_lincs_csv(url)`

Helper function to turn csv rows into dicts.

### 4.3.11 MeSH client (`indra.databases.mesh_client`)

`indra.databases.mesh_client.get_mesh_id_name(mesh_term, offline=False)`

Get the MESH ID and name for the given MESH term.

Uses the mappings table in `indra/resources`; if the MESH term is not listed there, falls back on the NLM REST API.

**Parameters**

- **mesh\_term** (*str*) – MESH Descriptor or Concept name, e.g. ‘Breast Cancer’.
- **offline** (*bool*) – Whether to allow queries to the NLM REST API if the given MESH term is not contained in INDRA’s internal MESH mappings file. Default is False (allows REST API queries).

**Returns** Returns a 2-tuple of the form (*id*, *name*) with the ID of the descriptor corresponding to the MESH label, and the descriptor name (which may not exactly match the name provided as an argument if it is a Concept name). If the query failed, or no descriptor corresponding to the name was found, returns a tuple of (None, None).

**Return type** tuple of strs

`indra.databases.mesh_client.get_mesh_id_name_from_web(mesh_term)`

Get the MESH ID and name for the given MESH term using the NLM REST API.

**Parameters** `mesh_term` (*str*) – MESH Descriptor or Concept name, e.g. ‘Breast Cancer’.

**Returns** Returns a 2-tuple of the form (*id, name*) with the ID of the descriptor corresponding to the MESH label, and the descriptor name (which may not exactly match the name provided as an argument if it is a Concept name). If the query failed, or no descriptor corresponding to the name was found, returns a tuple of (None, None).

**Return type** tuple of strs

`indra.databases.mesh_client.get_mesh_name(mesh_id, offline=False)`

Get the MESH label for the given MESH ID.

Uses the mappings table in *indra/resources*; if the MESH ID is not listed there, falls back on the NLM REST API.

**Parameters**

- **mesh\_id** (*str*) – MESH Identifier, e.g. ‘D003094’.
- **offline** (*bool*) – Whether to allow queries to the NLM REST API if the given MESH ID is not contained in INDRA’s internal MESH mappings file. Default is False (allows REST API queries).

**Returns** Label for the MESH ID, or None if the query failed or no label was found.

**Return type** str

`indra.databases.mesh_client.get_mesh_name_from_web`

Get the MESH label for the given MESH ID using the NLM REST API.

**Parameters** `mesh_id` (*str*) – MESH Identifier, e.g. ‘D003094’.

**Returns** Label for the MESH ID, or None if the query failed or no label was found.

**Return type** str

### 4.3.12 GO client (`indra.databases.go_client`)

`indra.databases.go_client.get_go_id_from_label(label)`

Get ID corresponding to a given GO label.

**Parameters** `label` (*str*) – The GO label to get the ID for.

**Returns** Identifier corresponding to the GO label, starts with GO:.

**Return type** str

`indra.databases.go_client.get_go_label(go_id)`

Get label corresponding to a given GO identifier.

**Parameters** `go_id` (*str*) – The GO identifier. Should include the *GO:* prefix, e.g., *GO:1903793* (positive regulation of anion transport).

**Returns** Label corresponding to the GO ID.

**Return type** str

`indra.databases.go_client.load_go_graph(go_fname)`

Load the GO data from an OWL file and parse into an RDF graph.

**Parameters** `go_fname` (*str*) – Path to the GO OWL file. Can be downloaded from <http://geneontology.org/ontology/go.owl>.

**Returns** RDF graph containing GO data.

**Return type** `rdflib.Graph`

`indra.databases.go_client.update_id_mappings(g)`

Compile all ID->label mappings and save to a TSV file.

**Parameters** `g` (*rdflib.Graph*) – RDF graph containing GO data.

`indra.databases.go_client.update_secondary_mappings(g)`

Compile all secondary ID->primary ID mappings and save to a TSV file.

**Parameters** `g` (*rdflib.Graph*) – RDF graph containing GO data.

#### 4.3.13 PubChem client (`indra.databases.pubchem_client`)

`indra.databases.pubchem_client.get_inchi_key`

Return the InChIKey for a given PubChem CID.

**Parameters** `pubchem_cid` (*str*) – The PubChem CID whose InChIKey should be returned.

**Returns** The InChIKey corresponding to the PubChem CID.

**Return type** `str`

#### 4.3.14 miRBase client (`indra.databases.mirbase_client`)

A client to miRBase.

`indra.databases.mirbase_client.get_mirbase_id_from_mirbase_name(mirbase_name)`

Return the miRBase identifier corresponding to the given miRBase name.

**Parameters** `mirbase_name` (*str*) – The miRBase ID to be converted. Example: “hsa-mir-19b-2”

**Returns** `mirbase_id` – The miRBase ID corresponding to the given miRBase name.

**Return type** `str`

`indra.databases.mirbase_client.get_mirbase_name_from_mirbase_id(mirbase_id)`

Return the miRBase name corresponding to the given miRBase ID.

**Parameters** `mirbase_id` (*str*) – The miRBase ID to be converted. Example: “MI0000060”

**Returns** `mirbase_name` – The miRBase name corresponding to the given miRBase ID.

**Return type** `str`

`indra.databases.mirbase_client.get_hgnc_id_from_mirbase_id(mirbase_id)`

Return the HGNC ID corresponding to the given miRBase ID.

**Parameters** `mirbase_id` (*str*) – The miRBase ID to be converted. Example: “MI0000060”

**Returns** `hgnc_id` – The HGNC ID corresponding to the given miRBase ID.

**Return type** `str`

`indra.databases.mirbase_client.get_mirbase_id_from_hgnc_id(hgnc_id)`

Return the miRBase ID corresponding to the given HGNC ID.

**Parameters** `hgnc_id` (*str*) – An HGNC identifier to convert to miRBase, if it is indeed an miRNA. Example: “31476”

**Returns** `mirbase_id` – The miRBase ID corresponding to the given HGNC ID.

**Return type** `str`

`indra.databases.mirbase_client.get_mirbase_id_from_hgnc_symbol` (*hgnc\_symbol*)

Return the HGNC gene symbol corresponding to the given miRBase ID.

**Parameters** `hgnc_symbol` (*str*) – An HGNC gene symbol to convert to miRBase, if it is indeed an miRNA. Example: “MIR19B2”

**Returns** `mirbase_id` – The miRBase ID corresponding to the given HGNC gene symbol.

**Return type** `str`

#### 4.3.15 Experimental Factor Ontology (EFO) client (`indra.databases.efo_client`)

A client to EFO.

`indra.databases.efo_client.get_efo_id_from_efo_name` (*efo\_name*)

Return the EFO identifier corresponding to the given EFO name.

**Parameters** `efo_name` (*str*) – The EFO name to be converted. Example: “gum cancer”

**Returns** `efo_id` – The EFO identifier corresponding to the given EFO name.

**Return type** `str`

`indra.databases.efo_client.get_efo_name_from_efo_id` (*efo\_id*)

Return the EFO name corresponding to the given EFO ID.

**Parameters** `efo_id` (*str*) – The EFO identifier to be converted. Example: “0005557”

**Returns** `efo_name` – The EFO name corresponding to the given EFO identifier.

**Return type** `str`

#### 4.3.16 Human Phenotype Ontology (HP) client (`indra.databases.hp_client`)

A client to HP.

`indra.databases.hp_client.get_hp_id_from_hp_name` (*hp\_name*)

Return the HP identifier corresponding to the given HP name.

**Parameters** `hp_name` (*str*) – The HP name to be converted. Example: “Nocturia”

**Returns** `hp_id` – The HP identifier corresponding to the given HP name.

**Return type** `str`

`indra.databases.hp_client.get_hp_name_from_hp_id` (*hp\_id*)

Return the HP name corresponding to the given HP ID.

**Parameters** `hp_id` (*str*) – The HP identifier to be converted. Example: “HP:0000017”

**Returns** `hp_name` – The HP name corresponding to the given HP identifier.

**Return type** `str`

### 4.3.17 Disease Ontology (DOID) client (`indra.databases.doid_client`)

A client to the Disease Ontology.

`indra.databases.doid_client.get_doid_id_from_doid_alt_id(doid_alt_id)`

Return the identifier corresponding to the given Disease Ontology alt id.

**Parameters** `doid_alt_id` (*str*) – The Disease Ontology alt id to be converted. Example: “DOID:267”

**Returns** `doid_id` – The Disease Ontology identifier corresponding to the given alt id.

**Return type** `str`

`indra.databases.doid_client.get_doid_id_from_doid_name(doid_name)`

Return the identifier corresponding to the given Disease Ontology name.

**Parameters** `doid_name` (*str*) – The Disease Ontology name to be converted. Example: “Nocturia”

**Returns** `doid_id` – The Disease Ontology identifier corresponding to the given name.

**Return type** `str`

`indra.databases.doid_client.get_doid_name_from_doid_id(doid_id)`

Return the name corresponding to the given Disease Ontology ID.

**Parameters** `doid_id` (*str*) – The Disease Ontology identifier to be converted. Example: “DOID:0000017”

**Returns** `doid_name` – The DOID name corresponding to the given DOID identifier.

**Return type** `str`

## 4.4 Literature clients (`indra.literature`)

`indra.literature.get_full_text(paper_id, idtype, preferred_content_type='text/xml')`

Return the content and the content type of an article.

This function retrieves the content of an article by its PubMed ID, PubMed Central ID, or DOI. It prioritizes full text content when available and returns an abstract from PubMed as a fallback.

#### Parameters

- **paper\_id** (*string*) – ID of the article.
- **idtype** (*'pmid', 'pmcid', or 'doi'*) – Type of the ID.
- **preferred\_content\_type** (*Optional[*str*]*) – Preference for full-text format, if available. Can be one of ‘text/xml’, ‘text/plain’, ‘application/pdf’. Default: ‘text/xml’

#### Returns

- **content** (*str*) – The content of the article.
- **content\_type** (*str*) – The content type of the article

`indra.literature.id_lookup(paper_id, idtype)`

Take an ID of type PMID, PMCID, or DOI and lookup the other IDs.

If the DOI is not found in Pubmed, try to obtain the DOI by doing a reverse-lookup of the DOI in CrossRef using article metadata.

#### Parameters

- **paper\_id** (*str*) – ID of the article.
- **idtype** (*str*) – Type of the ID: ‘pmid’, ‘pmcid’, or ‘doi’

**Returns** **ids** – A dictionary with the following keys: pmid, pmcid and doi.

**Return type** dict

#### 4.4.1 Pubmed client (`indra.literature.pubmed_client`)

Search and get metadata for articles in Pubmed.

`indra.literature.pubmed_client.expand_pagination` (*pages*)  
Convert a page number to long form, e.g., from 456-7 to 456-457.

`indra.literature.pubmed_client.get_abstract` (*pubmed\_id*, *prepend\_title=True*)  
Get the abstract of an article in the Pubmed database.

`indra.literature.pubmed_client.get_article_xml`  
Get the XML metadata for a single article from the Pubmed database.

`indra.literature.pubmed_client.get_id_count` (*search\_term*)  
Get the number of citations in Pubmed for a search query.

**Parameters** **search\_term** (*str*) – A term for which the PubMed search should be performed.

**Returns** The number of citations for the query, or None if the query fails.

**Return type** int or None

`indra.literature.pubmed_client.get_ids`  
Search Pubmed for paper IDs given a search term.

Search options can be passed as keyword arguments, some of which are custom keywords identified by this function, while others are passed on as parameters for the request to the PubMed web service For details on parameters that can be used in PubMed searches, see <https://www.ncbi.nlm.nih.gov/books/NBK25499/#chapter4.ESearch> Some useful parameters to pass are `db='pmc'` to search PMC instead of pubmed `reldate=2` to search for papers within the last 2 days `mindate='2016/03/01'`, `maxdate='2016/03/31'` to search for papers in March 2016.

PubMed, by default, limits returned PMIDs to a small number, and this number can be controlled by the “ret-max” parameter. This function uses a `retmax` value of 100,000 by default that can be changed via the corresponding keyword argument.

##### Parameters

- **search\_term** (*str*) – A term for which the PubMed search should be performed.
- **use\_text\_word** (*Optional[bool]*) – If True, the “[tw]” string is appended to the search term to constrain the search to “text words”, that is words that appear as whole in relevant parts of the PubMed entry (excl. for instance the journal name or publication date) like the title and abstract. Using this option can eliminate spurious search results such as all articles published in June for a search for the “JUN” gene, or journal names that contain Acad for a search for the “ACAD” gene. See also: [https://www.nlm.nih.gov/bsd/disted/pubmedtutorial/020\\_760.html](https://www.nlm.nih.gov/bsd/disted/pubmedtutorial/020_760.html) Default : True
- **kwargs** (*kwargs*) – Additional keyword arguments to pass to the PubMed search as parameters.

`indra.literature.pubmed_client.get_ids_for_gene`  
Get the curated set of articles for a gene in the Entrez database.

Search parameters for the Gene database query can be passed in as keyword arguments.



**Parameters** `hgnc_name` (*str*) – The HGNC name of the gene. This is used to obtain the HGNC ID (using the `hgnc_client` module) and in turn used to obtain the Entrez ID associated with the gene. Entrez is then queried for that ID.

```
indra.literature.pubmed_client.get_ids_for_mesh(mesh_id,          major_topic=False,
                                               **kwargs)
```

Return PMIDs that are annotated with a given MeSH ID.

#### Parameters

- `mesh_id` (*str*) – The MeSH ID of a term to search for, e.g., D009101.
- `major_topic` (*bool*) – If True, only papers for which the given MeSH ID is annotated as a major topic are returned. Otherwise all annotations are considered. Default: False
- `**kwargs` – Any further PubMed search arguments that are passed to `get_ids`.

```
indra.literature.pubmed_client.get_issns_for_journal
```

Get a list of the ISSN numbers for a journal given its NLM ID.

Information on NLM XML DTDs is available at <https://www.nlm.nih.gov/databases/dtd/>

```
indra.literature.pubmed_client.get_metadata_for_ids(pmid_list,
                                                    get_issns_from_nlm=False,
                                                    get_abstracts=False,
                                                    prepend_title=False)
```

Get article metadata for up to 200 PMIDs from the Pubmed database.

#### Parameters

- `pmid_list` (*list of str*) – Can contain 1-200 PMIDs.
- `get_issns_from_nlm` (*bool*) – Look up the full list of ISSN number for the journal associated with the article, which helps to match articles to CrossRef search results. Defaults to False, since it slows down performance.
- `get_abstracts` (*bool*) – Indicates whether to include the Pubmed abstract in the results.
- `prepend_title` (*bool*) – If `get_abstracts` is True, specifies whether the article title should be prepended to the abstract text.

**Returns** Dictionary indexed by PMID. Each value is a dict containing the following fields: ‘doi’, ‘title’, ‘authors’, ‘journal\_title’, ‘journal\_abbrev’, ‘journal\_nlm\_id’, ‘issn\_list’, ‘page’.

**Return type** dict of dicts

```
indra.literature.pubmed_client.get_metadata_from_xml_tree(tree,
                                                         get_issns_from_nlm=False,
                                                         get_abstracts=False,
                                                         prepend_title=False,
                                                         mesh_annotations=False)
```

Get metadata for an XML tree containing PubmedArticle elements.

**Documentation on the XML structure can be found at:**

- [https://www.nlm.nih.gov/bsd/licensee/elements\\_descriptions.html](https://www.nlm.nih.gov/bsd/licensee/elements_descriptions.html)
- [https://www.nlm.nih.gov/bsd/licensee/elements\\_alphabetical.html](https://www.nlm.nih.gov/bsd/licensee/elements_alphabetical.html)

#### Parameters

- `tree` (*xml.etree.ElementTree*) – ElementTree containing one or more PubmedArticle elements.

- **get\_issns\_from\_nlm** (*bool*) – Look up the full list of ISSN number for the journal associated with the article, which helps to match articles to CrossRef search results. Defaults to False, since it slows down performance.
- **get\_abstracts** (*bool*) – Indicates whether to include the Pubmed abstract in the results.
- **prepend\_title** (*bool*) – If `get_abstracts` is True, specifies whether the article title should be prepended to the abstract text.
- **mesh\_annotations** (*bool*) – If True, extract mesh annotations from the pubmed entries and include in the returned data. If false, don't.

**Returns** Dictionary indexed by PMID. Each value is a dict containing the following fields: 'doi', 'title', 'authors', 'journal\_title', 'journal\_abbrev', 'journal\_nlm\_id', 'issn\_list', 'page'.

**Return type** dict of dicts

`indra.literature.pubmed_client.get_title(pubmed_id)`  
Get the title of an article in the Pubmed database.

#### 4.4.2 Pubmed Central client (`indra.literature.pmc_client`)

`indra.literature.pmc_client.extract_paragraphs(xml_string)`  
Returns list of paragraphs in an NLM XML.

**Parameters** `xml_string` (*str*) – String containing valid NLM XML.

**Returns** List of extracted paragraphs in an NLM XML

**Return type** list of str

`indra.literature.pmc_client.extract_text(xml_string)`  
Get text from the body of the given NLM XML string.

**Parameters** `xml_string` (*str*) – String containing valid NLM XML.

**Returns** Extracted plaintext.

**Return type** str

`indra.literature.pmc_client.filter_pmids(pmid_list, source_type)`  
Filter a list of PMIDs for ones with full text from PMC.

**Parameters**

- **pmid\_list** (*list of str*) – List of PMIDs to filter.
- **source\_type** (*string*) – One of 'fulltext', 'oa\_xml', 'oa\_txt', or 'auth\_xml'.

**Returns** PMIDs available in the specified source/format type.

**Return type** list of str

`indra.literature.pmc_client.get_xml(pmc_id)`  
Returns XML for the article corresponding to a PMC ID.

`indra.literature.pmc_client.id_lookup(paper_id, idtype=None)`

This function takes a Pubmed ID, Pubmed Central ID, or DOI and use the Pubmed ID mapping service and looks up all other IDs from one of these. The IDs are returned in a dictionary.

### 4.4.3 CrossRef client (`indra.literature.crossref_client`)

`indra.literature.crossref_client.doi_query` (*pmid*, *search\_limit=10*)

Get the DOI for a PMID by matching CrossRef and Pubmed metadata.

Searches CrossRef using the article title and then accepts search hits only if they have a matching journal ISSN and page number with what is obtained from the Pubmed database.

`indra.literature.crossref_client.get_fulltext_links` (*doi*)

Return a list of links to the full text of an article given its DOI. Each list entry is a dictionary with keys: - URL: the URL to the full text - content-type: e.g. text/xml or text/plain - content-version - intended-application: e.g. text-mining

`indra.literature.crossref_client.get_metadata`

Returns the metadata of an article given its DOI from CrossRef as a JSON dict

### 4.4.4 Elsevier client (`indra.literature.elsevier_client`)

For information on the Elsevier API, see:

- API Specification: [http://dev.elsevier.com/api\\_docs.html](http://dev.elsevier.com/api_docs.html)
- Authentication: [https://dev.elsevier.com/tecdoc\\_api\\_authentication.html](https://dev.elsevier.com/tecdoc_api_authentication.html)

`indra.literature.elsevier_client.check_entitlement` (*doi*)

Check whether IP and credentials enable access to content for a doi.

This function uses the entitlement endpoint of the Elsevier API to check whether an article is available to a given institution. Note that this feature of the API is itself not available for all institution keys.

`indra.literature.elsevier_client.download_article` (*id\_val*, *id\_type='doi'*,  
*on\_retry=False*)

Low level function to get an XML article for a particular id.

#### Parameters

- **id\_val** (*str*) – The value of the id.
- **id\_type** (*str*) – The type of id, such as pmid (a.k.a. pubmed\_id), doi, or eid.
- **on\_retry** (*bool*) – This function has a recursive retry feature, and this is the only time this parameter should be used.

**Returns content** – If found, the content string is returned, otherwise, None is returned.

**Return type** str or None

`indra.literature.elsevier_client.download_article_from_ids` (*\*\*id\_dict*)

Download an article in XML format from Elsevier matching the set of ids.

**Parameters** **<id\_type>** (*str*) – You can enter any combination of eid, doi, pmid, and/or pii. Ids will be checked in that order, until either content has been found or all ids have been checked.

**Returns content** – If found, the content is returned as a string, otherwise None is returned.

**Return type** str or None

`indra.literature.elsevier_client.download_from_search` (*query\_str*, *folder*,  
*do\_extract\_text=True*,  
*max\_results=None*)

Save raw text files based on a search for papers on ScienceDirect.

This performs a search to get PIIIs, downloads the XML corresponding to the PII, extracts the raw text and then saves the text into a file in the designated folder.

#### Parameters

- **query\_str** (*str*) – The query string to search with
- **folder** (*str*) – The local path to an existing folder in which the text files will be dumped
- **do\_extract\_text** (*bool*) – Choose whether to extract text from the xml, or simply save the raw xml files. Default is True, so text is extracted.
- **max\_results** (*int or None*) – Default is None. If specified, limit the number of results to the given maximum.

```
indra.literature.elsevier_client.extract_paragraphs(xml_string)
```

Get paragraphs from the body of the given Elsevier xml.

```
indra.literature.elsevier_client.extract_text(xml_string)
```

Get text from the body of the given Elsevier xml.

```
indra.literature.elsevier_client.get_abstract(doi)
```

Get the abstract text of an article from Elsevier given a doi.

```
indra.literature.elsevier_client.get_article(doi, output_format='txt')
```

Get the full body of an article from Elsevier.

#### Parameters

- **doi** (*str*) – The doi for the desired article.
- **output\_format** (*'txt' or 'xml'*) – The desired format for the output. Selecting 'txt' (default) strips all xml tags and joins the pieces of text in the main text, while 'xml' simply takes the tag containing the body of the article and returns it as is. In the latter case, downstream code needs to be able to interpret Elsevier's XML format.

**Returns content** – Either text content or xml, as described above, for the given doi.

**Return type** str

```
indra.literature.elsevier_client.get_dois
```

Search ScienceDirect through the API for articles and return DOIs.

#### Parameters

- **query\_str** (*str*) – The query string to search with.
- **year** (*Optional[str]*) – The year to constrain the search to.
- **loaded\_after** (*Optional[str]*) – Date formatted as 'yyyy-MM-dd'T'HH:mm:ssX' to constrain the search to articles loaded after this date. Example: 2019-06-01T00:00:00Z

**Returns dois** – The list of DOIs identifying the papers returned by the search.

**Return type** list[str]

```
indra.literature.elsevier_client.get_piis(query_str)
```

Search ScienceDirect through the API for articles and return PIIIs.

Note that ScienceDirect has a limitation in which a maximum of 6,000 PIIIs can be retrieved for a given search and therefore this call is internally broken up into multiple queries by a range of years and the results are combined.

**Parameters query\_str** (*str*) – The query string to search with

**Returns piis** – The list of PIIIs identifying the papers returned by the search

**Return type** list[str]

`indra.literature.elsevier_client.get_piis_for_date`  
Search ScienceDirect through the API for articles and return PIIIs.

**Parameters**

- **query\_str** (*str*) – The query string to search with.
- **year** (*Optional[str]*) – The year to constrain the search to.
- **loaded\_after** (*Optional[str]*) – Date formatted as ‘yyyy-MM-dd’T’HH:mm:ssX’ to constrain the search to articles loaded after this date. Example: 2019-06-01T00:00:00Z

**Returns piis** – The list of PIIIs identifying the papers returned by the search.

**Return type** list[str]

`indra.literature.elsevier_client.search_science_direct`  
Search ScienceDirect for a given field with a query string.

Users can specify which field they are interested in and only values from that field will be returned. It is also possible to restrict the search either to a specific year of publication or to papers published after a specific date.

**Parameters**

- **query\_str** (*str*) – The query string to search with.
- **field\_name** (*str*) – A name of the field of interest to be returned. Accepted values are: authors, doi, loadDate, openAccess, pages, pii, publicationDate, sourceTitle, title, uri, volumeIssue.
- **year** (*Optional[str]*) – The year to constrain the search to.
- **loaded\_after** (*Optional[str]*) – Date formatted as ‘yyyy-MM-dd’T’HH:mm:ssX’ to constrain the search to articles loaded after this date.

**Returns all\_parts** – The list of values from the field of interest identifying the papers returned by the search.

**Return type** list[str]

#### 4.4.5 NewsAPI client (`indra.literature.newsapi_client`)

This module provides a client for the NewsAPI web service (<https://newsapi.org/>). The web service requires an API key which is available after registering at <https://newsapi.org/account>. This key can be set as NEWSAPI\_API\_KEY in the INDRA config file or as an environmental variable with the same name.

NewsAPI also requires attribution e.g. “powered by NewsAPI.org” for derived uses.

`indra.literature.newsapi_client.send_request` (*endpoint, \*\*kwargs*)  
Return the response to a query as JSON from the NewsAPI web service.

The basic API is limited to 100 results which is chosen unless explicitly given as an argument. Beyond that, paging is supported through the “page” argument, if needed.

**Parameters**

- **endpoint** (*str*) – Endpoint to query, e.g. “everything” or “top-headlines”
- **kwargs** (*dict*) – A list of keyword arguments passed as parameters with the query. The basic ones are “q” which is the search query, “from” is a start date formatted as for instance 2018-06-10 and “to” is an end date with the same format.

**Returns res\_json** – The response from the web service as a JSON dict.

**Return type** dict

#### 4.4.6 Deft Tools (`indra.literature.adeft_tools`)

This file provides several functions helpful for acquiring texts for Adept disambiguation.

It offers the ability to get text content for articles containing a particular gene. This is useful for acquiring training texts for genes that do not appear in a defining pattern with a problematic shortform.

General XML processing is also provided that allows for extracting text from a source that may be either of Elsevier XML, NLM XML or raw text. This is helpful because it avoids having to know in advance the source of text content from the database.

`indra.literature.adeft_tools.filter_paragraphs` (*paragraphs*, *contains=None*)

Filter paragraphs to only those containing one of a list of strings

**Parameters**

- **paragraphs** (*list of str*) – List of plaintext paragraphs from an article
- **contains** (*str or list of str*) – Exclude paragraphs not containing this string as a token, or at least one of the strings in contains if it is a list

**Returns** Plaintext consisting of all input paragraphs containing at least one of the supplied tokens.

**Return type** str

`indra.literature.adeft_tools.get_text_content_for_gene` (*hgnc\_name*)

Get articles that have been annotated to contain gene in entrez

**Parameters** **hgnc\_name** (*str*) – HGNC name for gene

**Returns** **text\_content** – xmls of fulltext if available otherwise abstracts for all articles that haven been annotated in entrez to contain the given gene

**Return type** list of str

`indra.literature.adeft_tools.get_text_content_for_pmids` (*pmids*)

Get text content for articles given a list of their pmids

**Parameters** **pmids** (*list of str*) –

**Returns** **text\_content**

**Return type** list of str

`indra.literature.adeft_tools.universal_extract_paragraphs` (*xml*)

Extract paragraphs from xml that could be from different sources

First try to parse the xml as if it came from elsevier. if we do not have valid elsevier xml this will throw an exception. the text extraction function in the pmc client may not throw an exception when parsing elsevier xml, silently processing the xml incorrectly

**Parameters** **xml** (*str*) – Either an NLM xml, Elsevier xml or plaintext

**Returns** **paragraphs** – Extracted plaintext paragraphs from NLM or Elsevier XML

**Return type** str

`indra.literature.adeft_tools.universal_extract_text` (*xml*, *contains=None*)

Extract plaintext from xml that could be from different sources

**Parameters**

- **xml** (*str*) – Either an NLM xml, Elsevier xml, or plaintext

- **contains** (*str or list of str*) – Exclude paragraphs not containing this string, or at least one of the strings in contains if it is a list

**Returns** The concatenation of all paragraphs in the input xml, excluding paragraphs not containing one of the tokens in the list contains. Paragraphs are separated by new lines.

**Return type** str

## 4.5 Preassembly (`indra.preassembler`)

### 4.5.1 Preassembler (`indra.preassembler`)

**class** `indra.preassembler.Preassembler` (*hierarchies, stmts=None, matches\_fun=None, refinement\_fun=None*)

De-duplicates statements and arranges them in a specificity hierarchy.

#### Parameters

- **hierarchies** (`dict[indra.preassembler.hierarchy_manager]`) – A dictionary of hierarchies with keys such as ‘entity’ (hierarchy of entities, primarily specifying relationships between genes and their families) and ‘modification’ pointing to HierarchyManagers
- **stmts** (list of `indra.statements.Statement` or `None`) – A set of statements to perform pre-assembly on. If `None`, statements should be added using the `add_statements()` method.
- **matches\_fun** (*Optional[function]*) – A function which takes a `Statement` object as argument and returns a string key that is used for duplicate recognition. If supplied, it overrides the use of the built-in `matches_key` method of each `Statement` being assembled.
- **refinement\_fun** (*Optional[function]*) – A function which takes two `Statement` objects and a `hierarchies` dict as an argument and returns `True` or `False`. If supplied, it overrides the built-in `refinement_of` method of each `Statement` being assembled.

#### **stmts**

Starting set of statements for preassembly.

**Type** list of `indra.statements.Statement`

#### **unique\_stmts**

Statements resulting from combining duplicates.

**Type** list of `indra.statements.Statement`

#### **related\_stmts**

Top-level statements after building the refinement hierarchy.

**Type** list of `indra.statements.Statement`

#### **hierarchies**

A dictionary of hierarchies with keys such as ‘entity’ and ‘modification’ pointing to `HierarchyManagers`

**Type** `dict[indra.preassembler.hierarchy_manager]`

#### **add\_statements** (*stmts*)

Add to the current list of statements.

**Parameters** **stmts** (list of `indra.statements.Statement`) – Statements to add to the current list.

**combine\_duplicate\_stmts** (*stmts*)

Combine evidence from duplicate Statements.

Statements are deemed to be duplicates if they have the same key returned by the *matches\_key()* method of the Statement class. This generally means that statements must be identical in terms of their arguments and can differ only in their associated *Evidence* objects.

This function keeps the first instance of each set of duplicate statements and merges the lists of Evidence from all of the other statements.

**Parameters** *stmts* (list of `indra.statements.Statement`) – Set of statements to de-duplicate.

**Returns** Unique statements with accumulated evidence across duplicates.

**Return type** list of `indra.statements.Statement`

**Examples**

De-duplicate and combine evidence for two statements differing only in their evidence lists:

```
>>> from indra.preassembler.hierarchy_manager import hierarchies
>>> map2k1 = Agent('MAP2K1')
>>> mapk1 = Agent('MAPK1')
>>> stmt1 = Phosphorylation(map2k1, mapk1, 'T', '185',
... evidence=[Evidence(text='evidence 1')])
>>> stmt2 = Phosphorylation(map2k1, mapk1, 'T', '185',
... evidence=[Evidence(text='evidence 2')])
>>> pa = Preassembler(hierarchies)
>>> uniq_stmts = pa.combine_duplicate_stmts([stmt1, stmt2])
>>> uniq_stmts
[Phosphorylation(MAP2K1(), MAPK1(), T, 185)]
>>> sorted([e.text for e in uniq_stmts[0].evidence]) # doctest:+IGNORE_UNICODE
['evidence 1', 'evidence 2']
```

**combine\_duplicates** ()

Combine duplicates among *stmts* and save result in *unique\_stmts*.

A wrapper around the method `combine_duplicate_stmts()`.

**combine\_related** (*return\_toplevel=True, poolsize=None, size\_cutoff=100*)

Connect related statements based on their refinement relationships.

This function takes as a starting point the unique statements (with duplicates removed) and returns a modified flat list of statements containing only those statements which do not represent a refinement of other existing statements. In other words, the more general versions of a given statement do not appear at the top level, but instead are listed in the *supports* field of the top-level statements.

If *unique\_stmts* has not been initialized with the de-duplicated statements, `combine_duplicates()` is called internally.

After this function is called the attribute *related\_stmts* is set as a side-effect.

The procedure for combining statements in this way involves a series of steps:

1. The statements are grouped by type (e.g., Phosphorylation) and each type is iterated over independently.
2. Statements of the same type are then grouped according to their Agents' entity hierarchy component identifiers. For instance, ERK, MAPK1 and MAPK3 are all in the same connected component in the entity hierarchy and therefore all Statements of the same type referencing these entities will be



grouped. This grouping assures that relations are only possible within Statement groups and not among groups. For two Statements to be in the same group at this step, the Statements must be the same type and the Agents at each position in the Agent lists must either be in the same hierarchy component, or if they are not in the hierarchy, must have identical `entity_matches_keys`. Statements with None in one of the Agent list positions are collected separately at this stage.

3. Statements with None at either the first or second position are iterated over. For a statement with a None as the first Agent, the second Agent is examined; then the Statement with None is added to all Statement groups with a corresponding component or `entity_matches_key` in the second position. The same procedure is performed for Statements with None at the second Agent position.
4. The statements within each group are then compared; if one statement represents a refinement of the other (as defined by the `refinement_off()` method implemented for the Statement), then the more refined statement is added to the `supports` field of the more general statement, and the more general statement is added to the `supported_by` field of the more refined statement.
5. A new flat list of statements is created that contains only those statements that have no `supports` entries (statements containing such entries are not eliminated, because they will be retrievable from the `supported_by` fields of other statements). This list is returned to the caller.

On multi-core machines, the algorithm can be parallelized by setting the `poolsize` argument to the desired number of worker processes. This feature is only available in Python > 3.4.

---

**Note:** Subfamily relationships must be consistent across arguments

For now, we require that merges can only occur if the *isa* relationships are all in the *same direction for all the agents* in a Statement. For example, the two statement groups: *RAF\_family -> MEK1* and *BRAF -> MEK\_family* would not be merged, since BRAF *isa* RAF\_family, but MEK\_family is not a MEK1. In the future this restriction could be revisited.

---

### Parameters

- **return\_toplevel** (*Optional[bool]*) – If True only the top level statements are returned. If False, all statements are returned. Default: True
- **poolsize** (*Optional[int]*) – The number of worker processes to use to parallelize the comparisons performed by the function. If None (default), no parallelization is performed. NOTE: Parallelization is only available on Python 3.4 and above.
- **size\_cutoff** (*Optional[int]*) – Groups with `size_cutoff` or more statements are sent to worker processes, while smaller groups are compared in the parent process. Default value is 100. Not relevant when parallelization is not used.

**Returns** The returned list contains Statements representing the more concrete/refined versions of the Statements involving particular entities. The attribute `related_stmts` is also set to this list. However, if `return_toplevel` is False then all statements are returned, irrespective of level of specificity. In this case the relationships between statements can be accessed via the `supports/supported_by` attributes.

**Return type** list of `indra.statement.Statement`

### Examples

A more general statement with no information about a Phosphorylation site is identified as supporting a more specific statement:

```

>>> from indra.preassembler.hierarchy_manager import hierarchies
>>> braf = Agent('BRAF')
>>> map2k1 = Agent('MAP2K1')
>>> st1 = Phosphorylation(braf, map2k1)
>>> st2 = Phosphorylation(braf, map2k1, residue='S')
>>> pa = Preassembler(hierarchies, [st1, st2])
>>> combined_stmts = pa.combine_related() # doctest:+ELLIPSIS
>>> combined_stmts
[Phosphorylation(BRAF(), MAP2K1(), S)]
>>> combined_stmts[0].supported_by
[Phosphorylation(BRAF(), MAP2K1())]
>>> combined_stmts[0].supported_by[0].supports
[Phosphorylation(BRAF(), MAP2K1(), S)]

```

**find\_contradicts()**

Return pairs of contradicting Statements.

**Returns** `contradicts` – A list of Statement pairs that are contradicting.

**Return type** `list(tuple(Statement, Statement))`

`indra.preassembler.flatten_evidence(stmts, collect_from=None)`

Add evidence from *supporting* stmts to evidence for *supported* stmts.

**Parameters**

- **stmts** (list of `indra.statements.Statement`) – A list of top-level statements with associated supporting statements resulting from building a statement hierarchy with `combine_related()`.
- **collect\_from** (*str* in ('supports', 'supported\_by')) – String indicating whether to collect and flatten evidence from the *supports* attribute of each statement or the *supported\_by* attribute. If not set, defaults to 'supported\_by'.

**Returns** `stmts` – Statement hierarchy identical to the one passed, but with the evidence lists for each statement now containing all of the evidence associated with the statements they are supported by.

**Return type** list of `indra.statements.Statement`

**Examples**

Flattening evidence adds the two pieces of evidence from the supporting statement to the evidence list of the top-level statement:

```

>>> from indra.preassembler.hierarchy_manager import hierarchies
>>> braf = Agent('BRAF')
>>> map2k1 = Agent('MAP2K1')
>>> st1 = Phosphorylation(braf, map2k1,
... evidence=[Evidence(text='foo'), Evidence(text='bar')])
>>> st2 = Phosphorylation(braf, map2k1, residue='S',
... evidence=[Evidence(text='baz'), Evidence(text='bak')])
>>> pa = Preassembler(hierarchies, [st1, st2])
>>> pa.combine_related() # doctest:+ELLIPSIS
[Phosphorylation(BRAF(), MAP2K1(), S)]
>>> [e.text for e in pa.related_stmts[0].evidence] # doctest:+IGNORE_UNICODE
['baz', 'bak']
>>> flattened = flatten_evidence(pa.related_stmts)

```

(continues on next page)

(continued from previous page)

```
>>> sorted([e.text for e in flattened[0].evidence]) # doctest:+IGNORE_UNICODE
['bak', 'bar', 'baz', 'foo']
```

`indra.preassembler.flatten_stmts` (*stmts*)

Return the full set of unique stmts in a pre-assembled stmt graph.

The flattened list of statements returned by this function can be compared to the original set of unique statements to make sure no statements have been lost during the preassembly process.

**Parameters** *stmts* (list of `indra.statements.Statement`) – A list of top-level statements with associated supporting statements resulting from building a statement hierarchy with `combine_related()`.

**Returns** *stmts* – List of all statements contained in the hierarchical statement graph.

**Return type** list of `indra.statements.Statement`

## Examples

Calling `combine_related()` on two statements results in one top-level statement; calling `flatten_stmts()` recovers both:

```
>>> from indra.preassembler.hierarchy_manager import hierarchies
>>> braf = Agent('BRAF')
>>> map2k1 = Agent('MAP2K1')
>>> st1 = Phosphorylation(braf, map2k1)
>>> st2 = Phosphorylation(braf, map2k1, residue='S')
>>> pa = Preassembler(hierarchies, [st1, st2])
>>> pa.combine_related() # doctest:+ELLIPSIS
[Phosphorylation(BRAF(), MAP2K1(), S)]
>>> flattened = flatten_stmts(pa.related_stmts)
>>> flattened.sort(key=lambda x: x.matches_key())
>>> flattened
[Phosphorylation(BRAF(), MAP2K1()), Phosphorylation(BRAF(), MAP2K1(), S)]
```

`indra.preassembler.render_stmt_graph` (*statements*, *reduce=True*, *english=False*, *rankdir=None*, *agent\_style=None*)

Render the statement hierarchy as a pygraphviz graph.

### Parameters

- **stmts** (list of `indra.statements.Statement`) – A list of top-level statements with associated supporting statements resulting from building a statement hierarchy with `combine_related()`.
- **reduce** (*bool*) – Whether to perform a transitive reduction of the edges in the graph. Default is True.
- **english** (*bool*) – If True, the statements in the graph are represented by their English-assembled equivalent; otherwise they are represented as text-formatted Statements.
- **rank\_dir** (*str or None*) – Argument to pass through to the pygraphviz *AGraph* constructor specifying graph layout direction. In particular, a value of 'LR' specifies a left-to-right direction. If None, the pygraphviz default is used.
- **agent\_style** (*dict or None*) – Dict of attributes specifying the visual properties of nodes. If None, the following default attributes are used:

```
agent_style = {'color': 'lightgray', 'style': 'filled',
              'fontname': 'arial'}
```

**Returns** Pygraphviz graph with nodes representing statements and edges pointing from supported statements to supported\_by statements.

**Return type** pygraphviz.AGraph

## Examples

Pattern for getting statements and rendering as a Graphviz graph:

```
>>> from indra.preassembler.hierarchy_manager import hierarchies
>>> braf = Agent('BRAF')
>>> map2k1 = Agent('MAP2K1')
>>> st1 = Phosphorylation(braf, map2k1)
>>> st2 = Phosphorylation(braf, map2k1, residue='S')
>>> pa = Preassembler(hierarchies, [st1, st2])
>>> pa.combine_related() # doctest:+ELLIPSIS
[Phosphorylation(BRAF(), MAP2K1(), S)]
>>> graph = render_stmt_graph(pa.related_stmts)
>>> graph.write('example_graph.dot') # To make the DOT file
>>> graph.draw('example_graph.png', prog='dot') # To make an image
```

Resulting graph:



## 4.5.2 Hierarchy manager (`indra.preassembler.hierarchy_manager`)

```
class indra.preassembler.hierarchy_manager.HierarchyManager (rdf_file=None,
                                                             build_closure=None,
                                                             uri_as_name=True)
```

Store hierarchical relationships between different types of entities.

Used to store, e.g., entity hierarchies (proteins and protein families) and modification hierarchies (serine phosphorylation vs. phosphorylation).

### Parameters

- **rdf\_file** (*string*) – Path to the RDF file containing the hierarchy.
- **build\_closure** (*bool or list or None*) – If True, the transitive closure of the hierarchy is generated up front to speed up processing. If a list, the entries in the list are namespaces for which a transitive closure should be built. Namespaces not listed are ignored and isa/partof lookups will not use the transitive closure. If False, no transitive closure is built. Default: True
- **uri\_as\_name** (*Optional[bool]*) – If True, entries are accessed directly by their URIs. If False entries are accessed by finding their name through the hasName relationship. Default: True

### graph

The RDF graph containing the hierarchy.

**Type** instance of *rdflib.Graph*

**build\_transitive\_closure** (*rel, tc\_set*)

Build a transitive closure for a given relation in a given dict.

**build\_transitive\_closures** ()

Build the transitive closures of the hierarchy.

This method constructs dictionaries which contain terms in the hierarchy as keys and either all the “isa+” or “partof+” related terms as values.

**directly\_or\_indirectly\_related** (*ns1, id1, ns2, id2, closure\_set, relation\_func*)

Return True if two entities have the specified relationship.

This relation is constructed possibly through multiple links connecting the two entities directly or indirectly.

#### Parameters

- **ns1** (*str*) – Namespace code for an entity.
- **id1** (*str*) – URI for an entity.
- **ns2** (*str*) – Namespace code for an entity.
- **id2** (*str*) – URI for an entity.
- **closure\_set** (*set*) – A set containing tuples of entities that have the specified relationship, directly or indirectly. Empty if this has not been precomputed.
- **relation\_func** (*function*) – Function with arguments (node, graph) that generates objects with some relationship with node on the given graph.

**Returns** True if t1 has the specified relationship with t2, either directly or through a series of intermediates; False otherwise.

**Return type** bool

**extend\_with** (*rdf\_file*)

Extend the RDF graph of this HierarchyManager with another RDF file.

**Parameters** **rdf\_file** (*str*) – An RDF file which is parsed such that the current graph and the graph described by the file are merged.

**find\_entity**

Get the entity that has the specified name (or synonym).

**Parameters** **x** (*string*) – Name or synonym for the target entity.

**get\_children** (*uri*)

Return all (not just immediate) children of a given entry.

**Parameters** **uri** (*str*) – The URI of the entry whose children are to be returned. See the `get_uri` method to construct this URI from a name space and id.

**get\_parents** (*uri, type='all'*)

Return parents of a given entry.

#### Parameters

- **uri** (*str*) – The URI of the entry whose parents are to be returned. See the `get_uri` method to construct this URI from a name space and id.
- **type** (*str*) – ‘all’: return all parents irrespective of level; ‘immediate’: return only the immediate parents; ‘top’: return only the highest level parents

**is\_opposite** (*ns1, id1, ns2, id2*)

Return True if two entities are in an “is\_opposite” relationship

**Parameters**

- **ns1** (*str*) – Namespace code for an entity.
- **id1** (*str*) – URI for an entity.
- **ns2** (*str*) – Namespace code for an entity.
- **id2** (*str*) – URI for an entity.

**Returns** True if t1 has an “is\_opposite” relationship with t2.

**Return type** bool

**isa** (*ns1, id1, ns2, id2*)

Return True if one entity has an “isa” relationship to another.

**Parameters**

- **ns1** (*str*) – Namespace code for an entity.
- **id1** (*string*) – URI for an entity.
- **ns2** (*str*) – Namespace code for an entity.
- **id2** (*str*) – URI for an entity.

**Returns** True if t1 has an “isa” relationship with t2, either directly or through a series of intermediates; False otherwise.

**Return type** bool

**isa\_or\_partof** (*ns1, id1, ns2, id2*)

Return True if two entities are in an “isa” or “partof” relationship

**Parameters**

- **ns1** (*str*) – Namespace code for an entity.
- **id1** (*str*) – URI for an entity.
- **ns2** (*str*) – Namespace code for an entity.
- **id2** (*str*) – URI for an entity.

**Returns** True if t1 has a “isa” or “partof” relationship with t2, either directly or through a series of intermediates; False otherwise.

**Return type** bool

**load\_from\_rdf\_file** (*rdf\_file*)

Initialize given an RDF input file representing the hierarchy.”

**Parameters** **rdf\_file** (*str*) – Path to an RDF file.

**load\_from\_rdf\_graph** (*rdf\_graph*)

Initialize given an RDF Graph representing the hierarchy.”

**Parameters** **rdf\_graph** (*rdflib.Graph*) – An rdflib Graph representing the hierarchy.

**load\_from\_rdf\_string** (*rdf\_str*)

Initialize given an RDF string representing the hierarchy.”

**Parameters** **rdf\_str** (*str*) – An RDF string.

**partof** (*ns1, id1, ns2, id2*)

Return True if one entity is “partof” another.

**Parameters**

- **ns1** (*str*) – Namespace code for an entity.
- **id1** (*str*) – URI for an entity.
- **ns2** (*str*) – Namespace code for an entity.
- **id2** (*str*) – URI for an entity.

**Returns** True if t1 has a “partof” relationship with t2, either directly or through a series of intermediates; False otherwise.

**Return type** bool

**exception** `indra.preassembler.hierarchy_manager.UnknownNamespaceException`

**class** `indra.preassembler.hierarchy_manager.YamlHierarchyManager` (*root*,  
*yaml\_to\_rdf*,  
*add\_leaves*)

Class to manage YAML-based hierarchies.

#### Parameters

- **root** (*dict*) – A YAML data structure loaded with the yamll package.
- **yaml\_to\_rdf** (*function*) – A function that takes the root object as an argument and returns an RDF graph.

**add\_entry** (*entry*, *examples=None*)

Add a given entry to the ontology.

#### Parameters

- **entry** (*str*) – An entry in the ontology, with parts separated by /, e.g., animals/mammals/dog.
- **examples** (*Optional[list[str]]*) – A list of strings that serve as examples for the given entry.

`indra.preassembler.hierarchy_manager.get_bio_hierarchies` (*from\_pickle=True*)

Return default hierarchies for the Bio context.

**Parameters** **from\_pickle** (*Optional[bool]*) – If True, hierarchies are loaded from a pre-generated pickle file. Otherwise, they are regenerated from RDF files (slower). Default: True

**Returns** A dict of hierarchy managers for each type of hierarchy.

**Return type** dict[str, *HierarchyManager*]

`indra.preassembler.hierarchy_manager.get_wm_hierarchies` ()

Return default hierarchy managers for the World Modeling context.

**Returns** A dict of hierarchy managers for each type of hierarchy, in this context only an *entity* hierarchy is provided in the dict.

**Return type** dict[str, *HierarchyManager*]

### 4.5.3 Entity grounding curation and mapping (`indra.preassembler.grounding_mapper`)

## Grounding mapping and standardization

```
class indra.preassembler.grounding_mapper.mapper.GroundingMapper (grounding_map=None,  
agent_map=None,  
ig-  
nores=None,  
misground-  
ing_map=None,  
use_adeft=True)
```

Maps grounding of INDRA Agents based on a given grounding map.

Each parameter, if not provided will result in loading the corresponding built-in grounding resource. To explicitly avoid loading the default, pass in an empty data structure as the given parameter, e.g., `ignores=[]`.

### Parameters

- **grounding\_map** (*Optional[dict]*) – The grounding map, a dictionary mapping strings (entity names) to a dictionary of database identifiers.
- **agent\_map** (*Optional[dict]*) – A dictionary mapping strings to grounded INDRA Agents with given state.
- **ignores** (*Optional[list]*) – A list of entity strings that, if encountered will result in the corresponding Statement being discarded.
- **misgrounding\_map** (*Optional[dict]*) – A mapping dict similar to the grounding map which maps entity strings to a given grounding which is known to be incorrect and should be removed if encountered (making the remaining Agent ungrounded).
- **use\_adeft** (*Optional[bool]*) – If True, Adept will be attempted to be used for disambiguation of acronyms. Default: True

**static check\_grounding\_map** (*gm*)

Run sanity checks on the grounding map, raise error if needed.

**map\_agent** (*agent, do\_rename*)

Return the given Agent with its grounding mapped.

This function grounds a single agent. It returns the new Agent object (which might be a different object if we load a new agent state from json) or the same object otherwise.

### Parameters

- **agent** (*indra.statements.Agent*) – The Agent to map.
- **do\_rename** (*bool*) – If True, the Agent name is updated based on the mapped grounding. If `do_rename` is True the priority for setting the name is FamPlex ID, HGNC symbol, then the gene name from Uniprot.

**Returns grounded\_agent** – The grounded Agent.

**Return type** *indra.statements.Agent*

**map\_agents\_for\_stmt** (*stmt, do\_rename=True*)

Return a new Statement whose agents have been grounding mapped.

### Parameters

- **stmt** (*indra.statements.Statement*) – The Statement whose agents need mapping.
- **do\_rename** (*Optional[bool]*) – If True, the Agent name is updated based on the mapped grounding. If `do_rename` is True the priority for setting the name is FamPlex ID, HGNC symbol, then the gene name from Uniprot. Default: True



**Returns** `mapped_stmt` – The mapped Statement.

**Return type** `indra.statements.Statement`

**map\_stmts** (*stmts*, *do\_rename=True*)

Return a new list of statements whose agents have been mapped

**Parameters**

- **stmts** (list of `indra.statements.Statement`) – The statements whose agents need mapping
- **do\_rename** (*Optional[bool]*) – If True, the Agent name is updated based on the mapped grounding. If `do_rename` is True the priority for setting the name is FamPlex ID, HGNC symbol, then the gene name from Uniprot. Default: True

**Returns** `mapped_stmts` – A list of statements given by mapping the agents from each statement in the input list

**Return type** list of `indra.statements.Statement`

**static rename\_agents** (*stmts*)

Return a list of mapped statements with updated agent names.

Creates a new list of statements without modifying the original list.

**Parameters** **stmts** (list of `indra.statements.Statement`) – List of statements whose Agents need their names updated.

**Returns** `mapped_stmts` – A new list of Statements with updated Agent names

**Return type** list of `indra.statements.Statement`

**static standardize\_agent\_name** (*agent*, *standardize\_refs=True*)

Standardize the name of an Agent based on grounding information.

If an agent contains a FamPlex grounding, the FamPlex ID is used as a name. Otherwise if it contains a Uniprot ID, an attempt is made to find the associated HGNC gene name. If one can be found it is used as the agent name and the associated HGNC ID is added as an entry to the `db_refs`. Similarly, CHEBI, MESH and GO IDs are used in this order of priority to assign a standardized name to the Agent. If no relevant IDs are found, the name is not changed.

**Parameters**

- **agent** (*indra.statements.Agent*) – An INDRA Agent whose name attribute should be standardized based on grounding information.
- **standardize\_refs** (*Optional[bool]*) – If True, this function assumes that the Agent's `db_refs` need to be standardized, e.g., HGNC mapped to UP. Default: True

**static standardize\_db\_refs** (*db\_refs*)

Return a standardized db refs dict for a given db refs dict.

**Parameters** **db\_refs** (*dict*) – A dict of db refs that may not be standardized, i.e., may be missing an available UP ID corresponding to an existing HGNC ID.

**Returns** The `db_refs` dict with standardized entries.

**Return type** dict

**update\_agent\_db\_refs** (*agent*, *db\_refs*, *do\_rename=True*)

Update `db_refs` of agent using the grounding map

If the grounding map is missing one of the HGNC symbol or Uniprot ID, attempts to reconstruct one from the other.

### Parameters

- **agent** (`indra.statements.Agent`) – The agent whose db\_refs will be updated
- **db\_refs** (`dict`) – The db\_refs so set for the agent.
- **do\_rename** (`Optional[bool]`) – If True, the Agent name is updated based on the mapped grounding. If do\_rename is True the priority for setting the name is FamPlex ID, HGNC symbol, then the gene name from Uniprot. Default: True

```
indra.preassembler.grounding_mapper.mapper.load_grounding_map(grounding_map_path,  
                                                             lineterminator=  
                                                             '\r\n',  
                                                             hgnc_symbols=True)
```

Return a grounding map dictionary loaded from a csv file.

In the file pointed to by `grounding_map_path`, the number of name\_space ID pairs can vary per row and commas are used to pad out entries containing fewer than the maximum amount of name spaces appearing in the file. Lines should be terminated with

both a carriage return and a new line by default.

Optionally, one can specify another csv file (pointed to by `ignore_path`) containing agent texts that are degenerate and should be filtered out.

It is important to note that this function assumes that the mapping file entries for the HGNC key are symbols not IDs. These symbols are converted to IDs upon loading here.

### Parameters

- **grounding\_map\_path** (`str`) – Path to csv file containing grounding map information. Rows of the file should be of the form `<agent_text>,<name_space_1>,<ID_1>,...<name_space_n>,<ID_n>`
- **lineterminator** (`Optional[str]`) – Line terminator used in input csv file. Default:
- **hgnc\_symbols** (`Optional[bool]`) – Set to True if the grounding map file contains HGNC symbols rather than IDs. In this case, the entries are replaced by IDs. Default: True

**Returns** `g_map` – The grounding map constructed from the given files.

**Return type** `dict`

## Analysis scripts for grounding

```
indra.preassembler.grounding_mapper.analysis.agent_texts(agents)
```

Return a list of all agent texts from a list of agents.

None values are associated to agents without agent texts

**Parameters** `agents` (list of `indra.statements.Agent`) –

**Returns** agent texts from input list of agents

**Return type** list of `str/None`

```
indra.preassembler.grounding_mapper.analysis.agent_texts_with_grounding(stmts)
```

Return agent text groundings in a list of statements with their counts

**Parameters** `stmts` (list of `indra.statements.Statement`) –

**Returns**

List of tuples of the form (text: `str`, ((name\_space: `str`, ID: `str`, count: `int`)...), total\_count: `int`)

Where the counts within the tuple of groundings give the number of times an agent with the given `agent_text` appears grounded with the particular name space and ID. The `total_count` gives the total number of times an agent with text appears in the list of statements.

**Return type** list of tuple

```
indra.preassembler.grounding_mapper.analysis.all_agents(stmts)
```

Return a list of all of the agents from a list of statements.

Only agents that are not None and have a TEXT entry are returned.

**Parameters** `stmts` (list of `indra.statements.Statement`) –

**Returns** `agents` – List of agents that appear in the input list of indra statements.

**Return type** list of `indra.statements.Agent`

```
indra.preassembler.grounding_mapper.analysis.get_agents_with_name(name,
                                                                    stmts)
```

Return all agents within a list of statements with a particular name.

```
indra.preassembler.grounding_mapper.analysis.get_sentences_for_agent(text,
                                                                    stmts,
                                                                    max_sentences=None)
```

Returns evidence sentences with a given agent text from a list of statements.

**Parameters**

- **text** (*str*) – An agent text
- **stmts** (list of `indra.statements.Statement`) – INDRA Statements to search in for evidence statements.
- **max\_sentences** (*Optional[int/None]*) – Cap on the number of evidence sentences to return. Default: None

**Returns** `sentences` – Evidence sentences from the list of statements containing the given agent text.

**Return type** list of str

```
indra.preassembler.grounding_mapper.analysis.protein_map_from_twg(twg)
```

Build map of entity texts to validate protein grounding.

Looks at the grounding of the entity texts extracted from the statements and finds proteins where there is grounding to a human protein that maps to an HGNC name that is an exact match to the entity text. Returns a dict that can be used to update/expand the grounding map.

**Parameters** `twg` (*list of tuple*) – list of tuples of the form output by `agent_texts_with_grounding`

**Returns** `protein_map` – dict keyed on agent text with associated values {'TEXT': `agent_text`, 'UP': `uniprot_id`}. Entries are for agent texts where the grounding map was able to find human protein grounded to this `agent_text` in Uniprot.

**Return type** dict

```
indra.preassembler.grounding_mapper.analysis.save_base_map(filename,
                                                            grouped_by_text)
```

Dump a list of agents along with groundings and counts into a csv file

**Parameters**

- **filename** (*str*) – Filepath for output file
- **grouped\_by\_text** (*list of tuple*) – List of tuples of the form output by `agent_texts_with_grounding`

`indra.preassembler.grounding_mapper.analysis.save_sentences` (*twg, stmts, filename, agent\_limit=300*)

Write evidence sentences for stmts with ungrounded agents to csv file.

#### Parameters

- **twg** (*list of tuple*) – list of tuples of ungrounded agent\_texts with counts of the number of times they are mentioned in the list of statements. Should be sorted in descending order by the counts. This is of the form output by the function ungrounded texts.
- **stmts** (list of `indra.statements.Statement`) –
- **filename** (*str*) – Path to output file
- **agent\_limit** (*Optional[int]*) – Number of agents to include in output file. Takes the top agents by count.

`indra.preassembler.grounding_mapper.analysis.ungrounded_texts` (*stmts*)

Return a list of all ungrounded entities ordered by number of mentions

**Parameters** **stmts** (list of `indra.statements.Statement`) –

**Returns** **ungroundc** – list of tuples of the form (text: str, count: int) sorted in descending order by count.

**Return type** list of tuple

## Adeft disambiguation functions

`indra.preassembler.grounding_mapper.adeft.run_adeft_disambiguation` (*stmt, agent, idx*)

Run Adeft disambiguation on an Agent in a given Statement.

This function looks at the evidence of the given Statement and attempts to look up the full paper or the abstract for the evidence. If both of those fail, the evidence sentence itself is used for disambiguation. The disambiguation model corresponding to the Agent text is then called, and the highest scoring returned grounding is set as the Agent's new grounding.

The Statement's annotations as well as the Agent are modified in place and no value is returned.

#### Parameters

- **stmt** (*indra.statements.Statement*) – An INDRA Statement in which the Agent to be disambiguated appears.
- **agent** (*indra.statements.Agent*) – The Agent (potentially grounding mapped) which we want to disambiguate in the context of the evidence of the given Statement.
- **idx** (*int*) – The index of the new Agent's position in the Statement's agent list (needed to set annotations correctly).

## 4.5.4 Site curation and mapping (`indra.preassembler.sitemapper`)

`class indra.preassembler.sitemapper.MappedStatement` (*original\_stmt, mapped\_mods, mapped\_stmt*)

Information about a Statement found to have invalid sites.

#### Parameters

- **original\_stmt** (*indra.statements.Statement*) – The statement prior to mapping.

- **mapped\_mods** (*list of MappedSite*) – A list of MappedSite objects.
- **mapped\_stmt** (`indra.statements.Statement`) – The statement after mapping. Note that if no information was found in the site map, it will be identical to the original statement.

```
class indra.preassembler.sitemapper.SiteMapper (site_map=None,      use_cache=False,
                                                cache_path=None,
                                                do_methionine_offset=True,
                                                do_orthology_mapping=True,
                                                do_isoform_mapping=True)
```

Use site information to fix modification sites in Statements.

This is a wrapper around the protmapper package's ProtMapper class and adds all the additional functionality to handle INDRA Statements and Agents.

### Parameters

- **site\_map** (dict (as returned by `load_site_map()`)) – A dict mapping tuples of the form (*gene*, *orig\_res*, *orig\_pos*) to a tuple of the form (*correct\_res*, *correct\_pos*, *comment*), where *gene* is the string name of the gene (canonicalized to HGNC); *orig\_res* and *orig\_pos* are the residue and position to be mapped; *correct\_res* and *correct\_pos* are the corrected residue and position, and *comment* is a string describing the reason for the mapping (species error, isoform error, wrong residue name, etc.).
- **use\_cache** (*Optional[bool]*) – If True, the SITEMAPPER\_CACHE\_PATH from the config (or environment) is loaded and cached mappings are read and written to the given path. Otherwise, no cache is used. Default: False
- **do\_methionine\_offset** (*boolean*) – Whether to check for off-by-one errors in site position (possibly) attributable to site numbering from mature proteins after cleavage of the initial methionine. If True, checks the reference sequence for a known modification at 1 site position greater than the given one; if there exists such a site, creates the mapping. Default is True.
- **do\_orthology\_mapping** (*boolean*) – Whether to check sequence positions for known modification sites in mouse or rat sequences (based on PhosphoSitePlus data). If a mouse/rat site is found that is linked to a site in the human reference sequence, a mapping is created. Default is True.
- **do\_isoform\_mapping** (*boolean*) – Whether to check sequence positions for known modifications in other human isoforms of the protein (based on PhosphoSitePlus data). If a site is found that is linked to a site in the human reference sequence, a mapping is created. Default is True.

### Examples

Fixing site errors on both the modification state of an agent (MAP2K1) and the target of a Phosphorylation statement (MAPK1):

```
>>> map2k1_phos = Agent('MAP2K1', db_refs={'UP':'Q02750'}, mods=[
... ModCondition('phosphorylation', 'S', '217'),
... ModCondition('phosphorylation', 'S', '221')])
>>> mapk1 = Agent('MAPK1', db_refs={'UP':'P28482'})
>>> stmt = Phosphorylation(map2k1_phos, mapk1, 'T', '183')
>>> (valid, mapped) = default_mapper.map_sites([stmt])
>>> valid
[]
```

(continues on next page)

(continued from previous page)

```

>>> mapped # doctest:+IGNORE_UNICODE
[
MappedStatement:
  original_stmt: Phosphorylation(MAP2K1(mods: (phosphorylation, S, 217),
↳(phosphorylation, S, 221)), MAPK1(), T, 183)
  mapped_mods: MappedSite(up_id='Q02750', error_code=None, valid=False, orig_
↳res='S', orig_pos='217', mapped_id='Q02750', mapped_res='S', mapped_pos='218',
↳description='off by one', gene_name='MAP2K1')
  MappedSite(up_id='Q02750', error_code=None, valid=False, orig_
↳res='S', orig_pos='221', mapped_id='Q02750', mapped_res='S', mapped_pos='222',
↳description='off by one', gene_name='MAP2K1')
  MappedSite(up_id='P28482', error_code=None, valid=False, orig_
↳res='T', orig_pos='183', mapped_id='P28482', mapped_res='T', mapped_pos='185',
↳description='INFERRED_MOUSE_SITE', gene_name='MAPK1')
  mapped_stmt: Phosphorylation(MAP2K1(mods: (phosphorylation, S, 218),
↳(phosphorylation, S, 222)), MAPK1(), T, 185)
]
>>> ms = mapped[0]
>>> ms.original_stmt
Phosphorylation(MAP2K1(mods: (phosphorylation, S, 217), (phosphorylation, S,
↳221)), MAPK1(), T, 183)
>>> ms.mapped_mods # doctest:+IGNORE_UNICODE
[MappedSite(up_id='Q02750', error_code=None, valid=False, orig_res='S', orig_pos=
↳'217', mapped_id='Q02750', mapped_res='S', mapped_pos='218', description='off
↳by one', gene_name='MAP2K1'), MappedSite(up_id='Q02750', error_code=None,
↳valid=False, orig_res='S', orig_pos='221', mapped_id='Q02750', mapped_res='S',
↳mapped_pos='222', description='off by one', gene_name='MAP2K1'), MappedSite(up_
↳id='P28482', error_code=None, valid=False, orig_res='T', orig_pos='183', mapped_
↳id='P28482', mapped_res='T', mapped_pos='185', description='INFERRED_MOUSE_SITE
↳', gene_name='MAPK1')]
>>> ms.mapped_stmt
Phosphorylation(MAP2K1(mods: (phosphorylation, S, 218), (phosphorylation, S,
↳222)), MAPK1(), T, 185)

```

**map\_sites** (*stmts*)

Check a set of statements for invalid modification sites.

Statements are checked against Uniprot reference sequences to determine if residues referred to by post-translational modifications exist at the given positions.

If there is nothing amiss with a statement (modifications on any of the agents, modifications made in the statement, etc.), then the statement goes into the list of valid statements. If there is a problem with the statement, the offending modifications are looked up in the site map (*site\_map*), and an instance of *MappedStatement* is added to the list of mapped statements.

**Parameters** *stmts* (list of *indra.statement.Statement*) – The statements to check for site errors.

**Returns** 2-tuple containing (*valid\_statements*, *mapped\_statements*). The first element of the tuple is a list of valid statements (*indra.statement.Statement*) that were not found to contain any site errors. The second element of the tuple is a list of mapped statements (*MappedStatement*) with information on the incorrect sites and corresponding statements with correctly mapped sites.

**Return type** tuple

### 4.5.5 Ontology mapping (`indra.preassembler.ontology_mapper`)

```
class indra.preassembler.ontology_mapper.OntologyMapper (statements,           map-
                                                    pings=None,           symmet-
                                                    ric=True, scored=False)
```

A class to map between ontologies in grounded arguments of Statements.

#### Parameters

- **statements** (*list[indra.statement.Statement]*) – A list of INDRA Statements to map
- **mappings** (*Optional[list[tuple]]*) – A list of tuples that map ontology entries to one another
- **symmetric** (*Optional[bool]*) – If True, the mappings are interpreted as symmetric and will be applied in both directions
- **scored** (*Optional[bool]*) – If True, the mappings are assumed to be scored and the scores are propagated into the mapped groundings. If False, the scores don't need to be provided in the mappings and even if they are, they are ignored. Default: False

```
map_statements ()
```

Run the ontology mapping on the statements.

### 4.6 Belief Engine (`indra.belief`)

```
class indra.belief.BayesianScorer (prior_counts, subtype_counts)
```

This is a belief scorer which assumes a Beta prior and a set of prior counts of correct and incorrect instances for a given source. It exposes an interface to take additional counts and update its probability parameters which can then be used to calculate beliefs on a set of Statements.

#### Parameters

- **prior\_counts** (*dict*) – A dictionary of counts of the form [pos, neg] for each source.
- **subtype\_counts** (*dict*) – A dictionary of counts of the form [pos, neg] for each subtype within a source.

```
update_counts (prior_counts, subtype_counts)
```

Update the internal counts based on given new counts.

#### Parameters

- **prior\_counts** (*dict*) – A dictionary of counts of the form [pos, neg] for each source.
- **subtype\_counts** (*dict*) – A dictionary of counts of the form [pos, neg] for each subtype within a source.

```
update_probs ()
```

Update the internal probability values given the counts.

```
class indra.belief.BeliefEngine (scorer=None, matches_fun=None)
```

Assigns beliefs to INDRA Statements based on supporting evidence.

#### scorer

A BeliefScorer object that computes the prior probability of a statement given its statement type and evidence. Must implement the `score_statement` method which takes Statements and computes the belief score of a statement, and the `check_prior_probs` method which takes a list of INDRA Statements and

verifies that the scorer has all the information it needs to score every statement in the list, and raises an exception if not.

Type *BeliefScorer*

**set\_hierarchy\_probs** (*statements*)

Sets hierarchical belief probabilities for INDRA Statements.

The Statements are assumed to be in a hierarchical relation graph with the supports and supported\_by attribute of each Statement object having been set. The hierarchical belief probability of each Statement is calculated based on its prior probability and the probabilities propagated from Statements supporting it in the hierarchy graph.

**Parameters statements** (*list[indra.statements.Statement]*) – A list of INDRA Statements whose belief scores are to be calculated. Each Statement object's belief attribute is updated by this function.

**set\_linked\_probs** (*linked\_statements*)

Sets the belief probabilities for a list of linked INDRA Statements.

The list of LinkedStatement objects is assumed to come from the MechanismLinker. The belief probability of the inferred Statement is assigned the joint probability of its source Statements.

**Parameters linked\_statements** (*list[indra.mechlinker.LinkedStatement]*) – A list of INDRA LinkedStatements whose belief scores are to be calculated. The belief attribute of the inferred Statement in the LinkedStatement object is updated by this function.

**set\_prior\_probs** (*statements*)

Sets the prior belief probabilities for a list of INDRA Statements.

The Statements are assumed to be de-duplicated. In other words, each Statement in the list passed to this function is assumed to have a list of Evidence objects that support it. The prior probability of each Statement is calculated based on the number of Evidences it has and their sources.

**Parameters statements** (*list[indra.statements.Statement]*) – A list of INDRA Statements whose belief scores are to be calculated. Each Statement object's belief attribute is updated by this function.

**class** `indra.belief.BeliefPackage` (*statement\_key, evidences*)

**evidences**

Alias for field number 1

**statement\_key**

Alias for field number 0

**class** `indra.belief.BeliefScorer`

Base class for a belief engine scorer, which computes the prior probability of a statement given its type and evidence.

To use with the belief engine, make a subclass with methods implemented.

**check\_prior\_probs** (*statements*)

Make sure the scorer has all the information needed to compute belief scores of each statement in the provided list, and raises an exception otherwise.

**Parameters statements** (*list<indra.statements.Statement>*) – List of statements to check

**score\_statement** (*st, extra\_evidence=None*)

Computes the prior belief probability for an INDRA Statement.



The Statement is assumed to be de-duplicated. In other words, the Statement is assumed to have a list of Evidence objects that supports it. The prior probability of the Statement is calculated based on the number of Evidences it has and their sources.

#### Parameters

- **st** (*indra.statements.Statement*) – An INDRA Statements whose belief scores are to be calculated.
- **extra\_evidence** (*list[indra.statements.Evidence]*) – A list of Evidences that are supporting the Statement (that aren't already included in the Statement's own evidence list).

**Returns** **belief\_score** – The computed prior probability for the statement

**Return type** float

**class** `indra.belief.SimpleScorer` (*prior\_probs=None, subtype\_probs=None*)

Computes the prior probability of a statement given its type and evidence.

#### Parameters

- **prior\_probs** (*dict[dict]*) – A dictionary of prior probabilities used to override/extend the default ones. There are two types of prior probabilities: `rand` and `syst` corresponding to random error and systematic error rate for each knowledge source. The `prior_probs` dictionary has the general structure `{'rand': {'s1': pr1, ..., 'sn': prn}, 'syst': {'s1': ps1, ..., 'sn': psn}}` where `'s1' ... 'sn'` are names of input sources and `pr1 ... prn` and `ps1 ... psn` are error probabilities. Examples: `{'rand': {'some_source': 0.1}}` sets the random error rate for `some_source` to 0.1; `{'rand': {''}}`
- **subtype\_probs** (*dict[dict]*) – A dictionary of random error probabilities for knowledge sources. When a subtype random error probability is not specified, will just use the overall type prior in `prior_probs`. If `None`, will only use the priors for each rule.

**check\_prior\_probs** (*statements*)

Throw Exception if BeliefEngine parameter is missing.

Make sure the scorer has all the information needed to compute belief scores of each statement in the provided list, and raises an exception otherwise.

**Parameters** **statements** (*list[indra.statements.Statement]*) – List of statements to check

**score\_evidence\_list** (*evidences*)

Return belief score given a list of supporting evidences.

**score\_statement** (*st, extra\_evidence=None*)

Computes the prior belief probability for an INDRA Statement.

The Statement is assumed to be de-duplicated. In other words, the Statement is assumed to have a list of Evidence objects that supports it. The prior probability of the Statement is calculated based on the number of Evidences it has and their sources.

#### Parameters

- **st** (*indra.statements.Statement*) – An INDRA Statements whose belief scores are to be calculated.
- **extra\_evidence** (*list[indra.statements.Evidence]*) – A list of Evidences that are supporting the Statement (that aren't already included in the Statement's own evidence list).

**Returns** **belief\_score** – The computed prior probability for the statement

**Return type** float

`indra.belief.evidence_random_noise_prior` (*evidence*, *type\_probs*, *subtype\_probs*)  
Determines the random-noise prior probability for this evidence.

If the evidence corresponds to a subtype, and that subtype has a curated prior noise probability, use that.

Otherwise, gives the random-noise prior for the overall rule type.

`indra.belief.sample_statements` (*stmts*, *seed=None*)  
Return statements sampled according to belief.

Statements are sampled independently according to their belief scores. For instance, a Statement with a belief score of 0.7 will end up in the returned Statement list with probability 0.7.

**Parameters**

- **stmts** (*list[indra.statements.Statement]*) – A list of INDRA Statements to sample.
- **seed** (*Optional[int]*) – A seed for the random number generator used for sampling.

**Returns** `new_stmts` – A list of INDRA Statements that were chosen by random sampling according to their respective belief scores.

**Return type** `list[indra.statements.Statement]`

`indra.belief.tag_evidence_subtype` (*evidence*)

Returns the type and subtype of an evidence object as a string, typically the extraction rule or database from which the statement was generated.

For biopax, this is just the database name.

**Parameters** **statement** (*indra.statements.Evidence*) – The statement which we wish to subtype

**Returns** **types** – A tuple with (type, subtype), both strings Returns (type, None) if the type of statement is not yet handled in this function.

**Return type** tuple

## 4.7 Mechanism Linker (`indra.mechlinker`)

**class** `indra.mechlinker.AgentState` (*agent*, *evidence=None*)

A class representing Agent state without identifying a specific Agent.

**bound\_conditions**

**Type** `list[indra.statements.BoundCondition]`

**mods**

**Type** `list[indra.statements.ModCondition]`

**mutations**

**Type** `list[indra.statements.Mutation]`

**location**

**Type** `indra.statements.location`

**apply\_to** (*agent*)

Apply this object's state to an Agent.

**Parameters** `agent` (*indra.statements.Agent*) – The agent to which the state should be applied

**class** `indra.mechlinker.BaseAgent` (*name*)  
Represents all activity types and active forms of an Agent.

**Parameters**

- **name** (*str*) – The name of the BaseAgent
- **activity\_types** (*list[str]*) – A list of activity types that the Agent has
- **active\_states** (*dict*) – A dict of activity types and their associated Agent states
- **activity\_reductions** (*dict*) – A dict of activity types and the type they are reduced to by inference.

**class** `indra.mechlinker.BaseAgentSet`  
Container for a set of BaseAgents.

This class wraps a dict of BaseAgent instance and can be used to get and set BaseAgents.

**get\_create\_base\_agent** (*agent*)  
Return BaseAgent from an Agent, creating it if needed.

**Parameters** `agent` (*indra.statements.Agent*) –

**Returns** `base_agent`

**Return type** *indra.mechlinker.BaseAgent*

**class** `indra.mechlinker.LinkedStatement` (*source\_stmts, inferred\_stmt*)  
A tuple containing a list of source Statements and an inferred Statement.

The list of source Statements are the basis for the inferred Statement.

**Parameters**

- **source\_stmts** (*list[indra.statements.Statement]*) – A list of source Statements
- **inferred\_stmts** (*indra.statements.Statement*) – A Statement that was inferred from the source Statements.

**class** `indra.mechlinker.MechLinker` (*stmts=None*)  
Rewrite the activation pattern of Statements and derive new Statements.

The mechanism linker (MechLinker) traverses a corpus of Statements and uses various inference steps to make the activity types and active forms consistent among Statements.

**add\_statements** (*stmts*)  
Add statements to the MechLinker.

**Parameters** `stmts` (*list[indra.statements.Statement]*) – A list of Statements to add.

**gather\_explicit\_activities** ()  
Aggregate all explicit activities and active forms of Agents.

This function iterates over `self.statements` and extracts explicitly stated activity types and active forms for Agents.

**gather\_implicit\_activities** ()  
Aggregate all implicit activities and active forms of Agents.

Iterate over `self.statements` and collect the implied activities and active forms of Agents that appear in the Statements.

Note that using this function to collect implied Agent activities can be risky. Assume, for instance, that a Statement from a reading system states that EGF bound to EGFR phosphorylates ERK. This would be interpreted as implicit evidence for the EGFR-bound form of EGF to have ‘kinase’ activity, which is clearly incorrect.

In contrast the alternative pair of this function: `gather_explicit_activities` collects only explicitly stated activities.

**static infer\_activations** (*stmts*)

Return inferred RegulateActivity from Modification + ActiveForm.

This function looks for combinations of Modification and ActiveForm Statements and infers Activation/Inhibition Statements from them. For example, if we know that A phosphorylates B, and the phosphorylated form of B is active, then we can infer that A activates B. This can also be viewed as having “explained” a given Activation/Inhibition Statement with a combination of more mechanistic Modification + ActiveForm Statements.

**Parameters** *stmts* (*list[indra.statements.Statement]*) – A list of Statements to infer RegulateActivity from.

**Returns** *linked\_stmts* – A list of LinkedStatements representing the inferred Statements.

**Return type** *list[indra.mechlinker.LinkedStatement]*

**static infer\_active\_forms** (*stmts*)

Return inferred ActiveForm from RegulateActivity + Modification.

This function looks for combinations of Activation/Inhibition Statements and Modification Statements, and infers an ActiveForm from them. For example, if we know that A activates B and A phosphorylates B, then we can infer that the phosphorylated form of B is active.

**Parameters** *stmts* (*list[indra.statements.Statement]*) – A list of Statements to infer ActiveForms from.

**Returns** *linked\_stmts* – A list of LinkedStatements representing the inferred Statements.

**Return type** *list[indra.mechlinker.LinkedStatement]*

**static infer\_complexes** (*stmts*)

Return inferred Complex from Statements implying physical interaction.

**Parameters** *stmts* (*list[indra.statements.Statement]*) – A list of Statements to infer Complexes from.

**Returns** *linked\_stmts* – A list of LinkedStatements representing the inferred Statements.

**Return type** *list[indra.mechlinker.LinkedStatement]*

**static infer\_modifications** (*stmts*)

Return inferred Modification from RegulateActivity + ActiveForm.

This function looks for combinations of Activation/Inhibition Statements and ActiveForm Statements that imply a Modification Statement. For example, if we know that A activates B, and phosphorylated B is active, then we can infer that A leads to the phosphorylation of B. An additional requirement when making this assumption is that the activity of B should only be dependent on the modified state and not other context - otherwise the inferred Modification is not necessarily warranted.

**Parameters** *stmts* (*list[indra.statements.Statement]*) – A list of Statements to infer Modifications from.

**Returns** *linked\_stmts* – A list of LinkedStatements representing the inferred Statements.

**Return type** list[indra.mechlinker.LinkedStatement]

**reduce\_activities** ()

Rewrite the activity types referenced in Statements for consistency.

Activity types are reduced to the most specific form whenever possible. For instance, if ‘kinase’ is the only specific activity type known for the BaseAgent of BRAF, its generic ‘activity’ forms are rewritten to ‘kinase’.

**replace\_activations** (linked\_stmts=None)

Remove RegulateActivity Statements that can be inferred out.

This function iterates over self.statements and looks for RegulateActivity Statements that either match or are refined by inferred RegulateActivity Statements that were linked (provided as the linked\_stmts argument). It removes RegulateActivity Statements from self.statements that can be explained by the linked statements.

**Parameters linked\_stmts** (Optional[list[indra.mechlinker.LinkedStatement]]) – A list of linked statements, optionally passed from outside. If None is passed, the MechLinker runs self.infer\_activations to infer RegulateActivities and obtain a list of LinkedStatements that are then used for removing existing Complexes in self.statements.

**replace\_complexes** (linked\_stmts=None)

Remove Complex Statements that can be inferred out.

This function iterates over self.statements and looks for Complex Statements that either match or are refined by inferred Complex Statements that were linked (provided as the linked\_stmts argument). It removes Complex Statements from self.statements that can be explained by the linked statements.

**Parameters linked\_stmts** (Optional[list[indra.mechlinker.LinkedStatement]]) – A list of linked statements, optionally passed from outside. If None is passed, the MechLinker runs self.infer\_complexes to infer Complexes and obtain a list of LinkedStatements that are then used for removing existing Complexes in self.statements.

**require\_active\_forms** ()

Rewrites Statements with Agents’ active forms in active positions.

As an example, the enzyme in a Modification Statement can be expected to be in an active state. Similarly, subjects of RegulateAmount and RegulateActivity Statements can be expected to be in an active form. This function takes the collected active states of Agents in their corresponding BaseAgents and then rewrites other Statements to apply the active Agent states to them.

**Returns new\_stmts** – A list of Statements which includes the newly rewritten Statements. This list is also set as the internal Statement list of the MechLinker.

**Return type** list[indra.statements.Statement]

## 4.8 Assemblers of model output (indra.assemblers)

### 4.8.1 Executable PySB models (indra.assemblers.pySB.assembler)

**PySB Assembler** (indra.assemblers.pySB.assembler)

**class** indra.assemblers.pySB.assembler.Param (name, value, unique=False)

Represent a parameter as an input to the assembly process.

**name**

The name of the parameter

**Type** str

**value**

The value of the parameter

**Type** float

**unique**

If True, a suffix is added to the end of the parameter name upon assembly to make sure the parameter is unique in the model. If False, the name attribute is used as is. Default: False

**Type** Optional[bool]

**class** `indra.assemblers.pysb.assembler.Policy` (*name, parameters=None, sites=None*)

Represent a policy that can be associated with a specific Statement.

**name**

The name of the policy, e.g. `one_step`

**Type** str

**parameters**

A dict of parameters where each key identifies the role of the parameter with respect to the policy, e.g. 'Km', and the value is a Param object.

**Type** dict[str, *Param*]

**sites**

A dict of site names corresponding to the interactions induced by the policy.

**Type** dict

**class** `indra.assemblers.pysb.assembler.PysbAssembler` (*statements=None*)

Assembler creating a PySB model from a set of INDRA Statements.

**Parameters** **statements** (*list[indra.statements.Statement]*) – A list of INDRA Statements to be assembled.

**policies**

A dictionary of policies that defines assembly policies for Statement types. It is assigned in the constructor.

**Type** dict

**statements**

A list of INDRA statements to be assembled.

**Type** list[indra.statements.Statement]

**model**

A PySB model object that is assembled by this class.

**Type** `pysb.Model`

**agent\_set**

A set of BaseAgents used during the assembly process.

**Type** *BaseAgentSet*

**add\_default\_initial\_conditions** (*value=None*)

Set default initial conditions in the PySB model.

**Parameters** **value** (*Optional[float]*) – Optionally a value can be supplied which will be the initial amount applied. Otherwise a built-in default is used.

**add\_statements** (*stmts*)

Add INDRA Statements to the assembler's list of statements.

**Parameters** *stmts* (*list[indra.statements.Statement]*) – A list of `indra.statements.Statement` to be added to the statement list of the assembler.

**export\_model** (*format, file\_name=None*)

Save the assembled model in a modeling formalism other than PySB.

For more details on exporting PySB models, see <http://pysb.readthedocs.io/en/latest/modules/export/index.html>

**Parameters**

- **format** (*str*) – The format to export into, for instance “kappa”, “bngl”, “sbml”, “matlab”, “mathematica”, “potterswheel”. See <http://pysb.readthedocs.io/en/latest/modules/export/index.html> for a list of supported formats. In addition to the formats supported by PySB itself, this method also provides “sbgn” output.
- **file\_name** (*Optional[str]*) – An optional file name to save the exported model into.

**Returns** *exp\_str* – The exported model string or object

**Return type** *str* or object

**make\_model** (*policies=None, initial\_conditions=True, reverse\_effects=False, model\_name='indra\_model'*)

Assemble the PySB model from the collected INDRA Statements.

This method assembles a PySB model from the set of INDRA Statements. The assembled model is both returned and set as the assembler's model argument.

**Parameters**

- **policies** (*Optional[Union[str, dict]]*) – A string or dictionary that defines one or more assembly policies.  
  
If *policies* is a string, it defines a global assembly policy that applies to all Statement types. Example: `one_step, interactions_only`  
  
A dictionary of policies has keys corresponding to Statement types and values to the policy to be applied to that type of Statement. For Statement types whose policy is undefined, the ‘default’ policy is applied. Example: `{‘Phosphorylation’: ‘two_step’}`
- **initial\_conditions** (*Optional[bool]*) – If True, default initial conditions are generated for the Monomers in the model. Default: True
- **reverse\_effects** (*Optional[bool]*) – If True, reverse rules are added to the model for activity, modification and amount regulations that have no corresponding reverse effects. Default: False
- **model\_name** (*Optional[str]*) – The name attribute assigned to the PySB Model object. Default: “indra\_model”

**Returns** *model* – The assembled PySB model object.

**Return type** `pysb.Model`

**print\_model** ()

Print the assembled model as a PySB program string.

This function is useful when the model needs to be passed as a string to another component.

**save\_model** (*file\_name='pysb\_model.py'*)

Save the assembled model as a PySB program file.

**Parameters** **file\_name** (*Optional[str]*) – The name of the file to save the model program code in. Default: pysb-model.py

**save\_rst** (*file\_name='pysb\_model.rst', module\_name='pysb\_module'*)

Save the assembled model as an RST file for literate modeling.

**Parameters**

- **file\_name** (*Optional[str]*) – The name of the file to save the RST in. Default: pysb\_model.rst
- **module\_name** (*Optional[str]*) – The name of the python function defining the module. Default: pysb\_module

**set\_context** (*cell\_type*)

Set protein expression amounts from CCLE as initial conditions.

This method uses `indra.databases.context_client` to get protein expression levels for a given cell type and set initial conditions for Monomers in the model accordingly.

**Parameters**

- **cell\_type** (*str*) – Cell type name for which expression levels are queried. The cell type name follows the CCLE database conventions.
- **Example** (*LOXIMVI\_SKIN, BT20\_BREAST*) –

**set\_expression** (*expression\_dict*)

Set protein expression amounts as initial conditions

**Parameters** **expression\_dict** (*dict*) – A dictionary in which the keys are gene names and the values are numbers representing the absolute amount (count per cell) of proteins expressed. Proteins that are not expressed can be represented as nan. Entries that are not in the dict or are in there but resolve to None, are set to the default initial amount. Example: {'EGFR': 12345, 'BRAF': 4567, 'ESR1': nan}

**exception** `indra.assemblers.pysb.assembler.UnknownPolicyError`

`indra.assemblers.pysb.assembler.add_rule_to_model` (*model, rule, annotations=None*)

Add a Rule to a PySB model and handle duplicate component errors.

`indra.assemblers.pysb.assembler.complex_monomers_default` (*stmt, agent\_set*)

In this (very simple) implementation, proteins in a complex are each given site names corresponding to each of the other members of the complex (lower case). So the resulting complex can be “fully connected” in that each member can be bound to all the others.

`indra.assemblers.pysb.assembler.complex_monomers_one_step` (*stmt, agent\_set*)

In this (very simple) implementation, proteins in a complex are each given site names corresponding to each of the other members of the complex (lower case). So the resulting complex can be “fully connected” in that each member can be bound to all the others.

`indra.assemblers.pysb.assembler.get_agent_rule_str` (*agent*)

Construct a string from an Agent as part of a PySB rule name.

`indra.assemblers.pysb.assembler.get_annotation` (*component, db\_name, db\_ref*)

Construct model Annotations for each component.

Annotation formats follow guidelines at <http://identifiers.org/>.

`indra.assemblers.pysb.assembler.get_create_parameter` (*model, param*)

Return parameter with given name, creating it if needed.



If `unique` is false and the parameter exists, the value is not changed; if it does not exist, it will be created. If `unique` is true then upon conflict a number is added to the end of the parameter name.

### Parameters

- **model** (*pysb.Model*) – The model to add the parameter to
- **param** (*Param*) – An assembly parameter object

```
indra.assemblers.pysb.assembler.get_monomer_pattern(model, agent, extra_fields=None)
```

Construct a PySB MonomerPattern from an Agent.

```
indra.assemblers.pysb.assembler.get_site_pattern(agent)
```

Construct a dictionary of Monomer site states from an Agent.

This crates the mapping to the associated PySB monomer from an INDRA Agent object.

```
indra.assemblers.pysb.assembler.get_uncond_agent(agent)
```

Construct the unconditional state of an Agent.

The unconditional Agent is a copy of the original agent but without any bound conditions and modification conditions. Mutation conditions, however, are preserved since they are static.

```
indra.assemblers.pysb.assembler.grounded_monomer_patterns(model, agent, ignore_activities=False)
```

Get monomer patterns for the agent accounting for grounding information.

### Parameters

- **model** (*pysb.core.Model*) – The model to search for MonomerPatterns matching the given Agent.
- **agent** (*indra.statements.Agent*) – The Agent to find matching MonomerPatterns for.
- **ignore\_activites** (*bool*) – Whether to ignore any ActivityConditions on the agent when determining the required site conditions for the MonomerPattern. For example, if set to True, will find a match for the agent *MAPK1(activity=kinase)* even if the corresponding MAPK1 Monomer in the model has no site named *kinase*. Default is False (more stringent matching).

### Returns

**Return type** generator of MonomerPatterns

```
indra.assemblers.pysb.assembler.parse_identifiers_url(url)
```

Parse an identifiers.org URL into (namespace, ID) tuple.

```
indra.assemblers.pysb.assembler.set_base_initial_condition(model, monomer, value)
```

Set an initial condition for a monomer in its ‘default’ state.

```
indra.assemblers.pysb.assembler.set_extended_initial_condition(model, monomer=None, value=0)
```

Set an initial condition for monomers in “modified” state.

This is useful when using downstream analysis that relies on reactions being active in the model. One example is BioNetGen-based reaction network diagram generation.

**PySB PreAssembler (indra.assemblers.pysb.preassembler)**

**class** indra.assemblers.pysb.preassembler.**PysbPreassembler** (*stmts=None*)  
Pre-assemble Statements in preparation for PySB assembly.

**Parameters** **stmts** (*list[indra.statements.Statement]*) – A list of Statements to assemble

**add\_reverse\_effects** ()

Add Statements for the reverse effects of some Statements.

For instance, if a protein is phosphorylated but never dephosphorylated in the model, we add a generic dephosphorylation here. This step is usually optional in the assembly process.

**add\_statements** (*stmts*)

Add a list of Statements for assembly.

**replace\_activities** ()

Replace active flags with Agent states when possible.

**Base Agents (indra.assemblers.pysb.base\_agents)**

**class** indra.assemblers.pysb.base\_agents.**BaseAgentSet**  
Container for a dict of BaseAgents with their names as keys.

**get\_create\_base\_agent** (*agent*)

Return base agent with given name, creating it if needed.

**items** ()

Return items for the set of BaseAgents that this class wraps.

**class** indra.assemblers.pysb.base\_agents.**BaseAgent** (*name*)  
A BaseAgent aggregates the global properties of an Agent.

The BaseAgent class aggregates the name, sites, site states, active forms, inactive forms and database references of Agents from individual INDRA Statements. This allows the PySB Assembler to correctly assemble the Monomer signatures in the model.

**add\_activity\_form** (*activity\_pattern, is\_active*)

Adds the pattern as an active or inactive form to an Agent.

**Parameters**

- **activity\_pattern** (*dict*) – A dictionary of site names and their states.
- **is\_active** (*bool*) – Is True if the given pattern corresponds to an active state.

**add\_activity\_type** (*activity\_type*)

Adds an activity type to an Agent.

**Parameters** **activity\_type** (*str*) – The type of activity to add such as ‘activity’, ‘kinase’, ‘gtpbound’

**add\_site\_states** (*site, states*)

Create new states on an agent site if the state doesn’t exist.

**create\_mod\_site** (*mc*)

Create modification site for the BaseAgent from a ModCondition.

**create\_site** (*site, states=None*)

Create a new site on an agent if it doesn’t already exist.

### A utility to get graphs from kappa (`indra.assemblers.pysb.kappa_util`)

`indra.assemblers.pysb.kappa_util.cm_json_to_graph(im_json)`

Return pygraphviz Agraph from Kappy's contact map JSON.

**Parameters** `im_json` (*dict*) – A JSON dict which contains a contact map generated by Kappy.

**Returns** `graph` – A graph representing the contact map.

**Return type** `pygraphviz.Agraph`

`indra.assemblers.pysb.kappa_util.im_json_to_graph(im_json)`

Return networkx graph from Kappy's influence map JSON.

**Parameters** `im_json` (*dict*) – A JSON dict which contains an influence map generated by Kappy.

**Returns** `graph` – A graph representing the influence map.

**Return type** `networkx.MultiDiGraph`

### 4.8.2 Cytoscape networks (`indra.assemblers.cx.assembler`)

**class** `indra.assemblers.cx.assembler.CxAssembler` (*stmts=None, network\_name=None*)

This class assembles a CX network from a set of INDRA Statements.

The CX format is an aspect oriented data mode for networks. The format is defined at <http://www.home.ndexbio.org/data-model/>. The CX format is the standard for NDEx and is compatible with CytoScape via the CyNDEx plugin.

#### Parameters

- **stmts** (*Optional[list[indra.statements.Statement]]*) – A list of INDRA Statements to be assembled.
- **network\_name** (*Optional[str]*) – The name of the network to be assembled. Default: `indra_assembled`

#### statements

A list of INDRA Statements to be assembled.

**Type** `list[indra.statements.Statement]`

#### network\_name

The name of the network to be assembled.

**Type** `str`

#### cx

The structure of the CX network that is assembled.

**Type** `dict`

#### add\_statements

 (*stmts*)

Add INDRA Statements to the assembler's list of statements.

**Parameters** `stmts` (*list[indra.statements.Statement]*) – A list of `indra.statements.Statement` to be added to the statement list of the assembler.

#### make\_model

 (*add\_indra\_json=True*)

Assemble the CX network from the collected INDRA Statements.

This method assembles a CX network from the set of INDRA Statements. The assembled network is set as the assembler's `cx` argument.

**Parameters** `add_indra_json` (*Optional[bool]*) – If True, the INDRA Statement JSON annotation is added to each edge in the network. Default: True

**Returns** `cx_str` – The json serialized CX model.

**Return type** str

**print\_cx** (*pretty=True*)

Return the assembled CX network as a json string.

**Parameters** `pretty` (*bool*) – If True, the CX string is formatted with indentation (for human viewing) otherwise no indentation is used.

**Returns** `json_str` – A json formatted string representation of the CX network.

**Return type** str

**save\_model** (*file\_name='model.cx'*)

Save the assembled CX network in a file.

**Parameters** `file_name` (*Optional[str]*) – The name of the file to save the CX network to. Default: model.cx

**set\_context** (*cell\_type*)

Set protein expression data and mutational status as node attribute

This method uses `indra.databases.context_client` to get protein expression levels and mutational status for a given cell type and set a node attribute for proteins accordingly.

**Parameters** `cell_type` (*str*) – Cell type name for which expression levels are queried. The cell type name follows the CCLE database conventions. Example: LOXIMVI\_SKIN, BT20\_BREAST

**upload\_model** (*ndex\_cred=None, private=True, style='default'*)

Creates a new NDEx network of the assembled CX model.

To upload the assembled CX model to NDEx, you need to have a registered account on NDEx (<http://ndexbio.org/>) and have the `ndex` python package installed. The uploaded network is private by default.

**Parameters**

- **ndex\_cred** (*Optional[dict]*) – A dictionary with the following entries: ‘user’: NDEx user name ‘password’: NDEx password
- **private** (*Optional[bool]*) – Whether or not the created network will be private on NDEX.
- **style** (*Optional[str]*) – This optional parameter can either be (1) The UUID of an existing NDEx network whose style should be applied to the new network. (2) Unspecified or ‘default’ to use the default INDRA-assembled network style. (3) None to not set a network style.

**Returns** `network_id` – The UUID of the NDEx network that was created by uploading the assembled CX model.

**Return type** str

### 4.8.3 Natural language (`indra.assemblers.english.assembler`)

**class** `indra.assemblers.english.assembler.EnglishAssembler` (*stmts=None*)

This assembler generates English sentences from INDRA Statements.

**Parameters** `stmts` (*Optional[list[indra.statements.Statement]]*) – A list of INDRA Statements to be added to the assembler.

**statements**

A list of INDRA Statements to assemble.

**Type** `list[indra.statements.Statement]`

**model**

The assembled sentences as a single string.

**Type** `str`

**add\_statements** (*stmts*)

Add INDRA Statements to the assembler’s list of statements.

**Parameters** `stmts` (*list[indra.statements.Statement]*) – A list of `indra.statements.Statement` to be added to the statement list of the assembler.

**make\_model** ()

Assemble text from the set of collected INDRA Statements.

**Returns** `stmt_strs` – Return the assembled text as unicode string. By default, the text is a single string consisting of one or more sentences with periods at the end.

**Return type** `str`

`indra.assemblers.english.assembler.english_join` (*lst*)

Join a list of strings according to English grammar.

**Parameters** `lst` (*list of str*) – A list of strings to join.

**Returns** A string which describes the list of elements, e.g., “apples, pears, and bananas”.

**Return type** `str`

`indra.assemblers.english.assembler.statement_base_verb` (*stmt\_type*)

Return the base verb form of a statement type.

**Parameters** `stmt_type` (*str*) – The lower case string form of a statement type, for instance, ‘phosphorylation’.

**Returns** The base verb form of a statement type, for instance, ‘phosphorylate’.

**Return type** `str`

`indra.assemblers.english.assembler.statement_passive_verb` (*stmt\_type*)

Return the passive / state verb form of a statement type.

**Parameters** `stmt_type` (*str*) – The lower case string form of a statement type, for instance, ‘phosphorylation’.

**Returns** The passive/state verb form of a statement type, for instance, ‘phosphorylated’.

**Return type** `str`

`indra.assemblers.english.assembler.statement_present_verb` (*stmt\_type*)

Return the present verb form of a statement type.

**Parameters** `stmt_type` (*str*) – The lower case string form of a statement type, for instance, ‘phosphorylation’.

**Returns** The present verb form of a statement type, for instance, ‘phosphorylates’.

**Return type** `str`

#### 4.8.4 Node-edge graphs (`indra.assemblers.graph.assembler`)

```
class indra.assemblers.graph.assembler.GraphAssembler (stmts=None,  
graph_properties=None,  
node_properties=None,  
edge_properties=None)
```

The Graph assembler assembles INDRA Statements into a Graphviz node-edge graph.

##### Parameters

- **stmts** (*Optional[list[indra.statements.Statement]]*) – A list of INDRA Statements to be added to the assembler’s list of Statements.
- **graph\_properties** (*Optional[dict[str: str]]*) – A dictionary of graphviz graph properties overriding the default ones.
- **node\_properties** (*Optional[dict[str: str]]*) – A dictionary of graphviz node properties overriding the default ones.
- **edge\_properties** (*Optional[dict[str: str]]*) – A dictionary of graphviz edge properties overriding the default ones.

##### statements

A list of INDRA Statements to be assembled.

**Type** list[indra.statements.Statement]

##### graph

A pygraphviz graph that is assembled by this assembler.

**Type** pygraphviz.AGraph

##### existing\_nodes

The list of nodes (identified by node key tuples) that are already in the graph.

**Type** list[tuple]

##### existing\_edges

The list of edges (identified by edge key tuples) that are already in the graph.

**Type** list[tuple]

##### graph\_properties

A dictionary of graphviz graph properties used for assembly.

**Type** dict[str: str]

##### node\_properties

A dictionary of graphviz node properties used for assembly.

**Type** dict[str: str]

##### edge\_properties

A dictionary of graphviz edge properties used for assembly. Note that most edge properties are determined based on the type of the edge by the assembler (e.g. color, arrowhead). These settings cannot be directly controlled through the API.

**Type** dict[str: str]

##### add\_statements (*stmts*)

Add a list of statements to be assembled.

**Parameters** **stmts** (*list[indra.statements.Statement]*) – A list of INDRA Statements to be appended to the assembler’s list.

**get\_string()**

Return the assembled graph as a string.

**Returns graph\_string** – The assembled graph as a string.

**Return type** str

**make\_model()**

Assemble the graph from the assembler's list of INDRA Statements.

**save\_dot** (*file\_name='graph.dot'*)

Save the graph in a graphviz dot file.

**Parameters file\_name** (*Optional[str]*) – The name of the file to save the graph dot string to.

**save\_pdf** (*file\_name='graph.pdf', prog='dot'*)

Draw the graph and save as an image or pdf file.

**Parameters**

- **file\_name** (*Optional[str]*) – The name of the file to save the graph as. Default: graph.pdf
- **prog** (*Optional[str]*) – The graphviz program to use for graph layout. Default: dot

#### 4.8.5 SIF / Boolean networks (`indra.assemblers.sif.assembler`)

**class** `indra.assemblers.sif.assembler.SifAssembler` (*stmts=None*)

The SIF assembler assembles INDRA Statements into a networkx graph.

This graph can then be exported into SIF (simple interaction format) or a Boolean network.

**Parameters stmts** (*Optional[list[indra.statements.Statement]]*) – A list of INDRA Statements to be added to the assembler's list of Statements.

**graph**

A networkx graph that is assembled by this assembler.

**Type** networkx.DiGraph

**make\_model** (*use\_name\_as\_key=False, include\_mods=False, include\_complexes=False*)

Assemble the graph from the assembler's list of INDRA Statements.

**Parameters**

- **use\_name\_as\_key** (*boolean*) – If True, uses the name of the agent as the key to the nodes in the network. If False (default) uses the `matches_key()` of the agent.
- **include\_mods** (*boolean*) – If True, adds Modification statements into the graph as directed edges. Default is False.
- **include\_complexes** (*boolean*) – If True, creates two edges (in both directions) between all pairs of nodes in Complex statements. Default is False.

**print\_boolean\_net** (*out\_file=None*)

Return a Boolean network from the assembled graph.

See <https://github.com/ialbert/booleannet> for details about the format used to encode the Boolean rules.

**Parameters out\_file** (*Optional[str]*) – A file name in which the Boolean network is saved.

**Returns full\_str** – The string representing the Boolean network.

**Return type** str

**print\_loopy** (*as\_url=True*)

Return

**Parameters** **out\_file** (*Optional[str]*) – A file name in which the Loopy network is saved.

**Returns** **full\_str** – The string representing the Loopy network.

**Return type** str

**print\_model** (*include\_unsigned\_edges=False*)

Return a SIF string of the assembled model.

**Parameters** **include\_unsigned\_edges** (*bool*) – If True, includes edges with an unknown activating/inactivating relationship (e.g., most PTMs). Default is False.

**save\_model** (*fname, include\_unsigned\_edges=False*)

Save the assembled model's SIF string into a file.

**Parameters**

- **fname** (*str*) – The name of the file to save the SIF into.
- **include\_unsigned\_edges** (*bool*) – If True, includes edges with an unknown activating/inactivating relationship (e.g., most PTMs). Default is False.

#### 4.8.6 MITRE “index cards” (`indra.assemblers.index_card.assembler`)

**class** `indra.assemblers.index_card.assembler.IndexCardAssembler` (*statements=None, pmc\_override=None*)

Assembler creating index cards from a set of INDRA Statements.

**Parameters**

- **statements** (*list*) – A list of INDRA statements to be assembled.
- **pmc\_override** (*Optional[str]*) – A PMC ID to assign to the index card.

**statements**

A list of INDRA statements to be assembled.

**Type** list

**add\_statements** (*statements*)

Add statements to the assembler.

**Parameters** **statements** (*list[indra.statement.Statements]*) – The list of Statements to add to the assembler.

**make\_model** ()

Assemble statements into index cards.

**print\_model** ()

Return the assembled cards as a JSON string.

**Returns** **cards\_json** – The JSON string representing the assembled cards.

**Return type** str

**save\_model** (*file\_name='index\_cards.json'*)

Save the assembled cards into a file.



**Parameters** `file_name` (*Optional[str]*) – The name of the file to save the cards into.  
Default: `index_cards.json`

#### 4.8.7 SBGN output (`indra.assemblers.sbgn.assembler`)

**class** `indra.assemblers.sbgn.assembler.SBGNAssembler` (*statements=None*)

This class assembles an SBGN model from a set of INDRA Statements.

The Systems Biology Graphical Notation (SBGN) is a widely used graphical notation standard for systems biology models. This assembler creates SBGN models following the Process Description (PD) standard, documented at: [https://github.com/sbgn/process-descriptions/blob/master/UserManual/sbgn\\_PD-level1-user-public.pdf](https://github.com/sbgn/process-descriptions/blob/master/UserManual/sbgn_PD-level1-user-public.pdf). For more information on SBGN, see: <http://sbgn.github.io/sbgn/>

**Parameters** `stmts` (*Optional[list[indra.statements.Statement]]*) – A list of INDRA Statements to be assembled.

**statements**

A list of INDRA Statements to be assembled.

**Type** `list[indra.statements.Statement]`

**sbgn**

The structure of the SBGN model that is assembled, represented as an XML ElementTree.

**Type** `lxml.etree.ElementTree`

**add\_statements** (*stmts*)

Add INDRA Statements to the assembler's list of statements.

**Parameters** `stmts` (*list[indra.statements.Statement]*) – A list of `indra.statements.Statement` to be added to the statement list of the assembler.

**make\_model** ()

Assemble the SBGN model from the collected INDRA Statements.

This method assembles an SBGN model from the set of INDRA Statements. The assembled model is set as the assembler's `sbgn` attribute (it is represented as an XML ElementTree internally). The model is returned as a serialized XML string.

**Returns** `sbgn_str` – The XML serialized SBGN model.

**Return type** `str`

**print\_model** (*pretty=True, encoding='utf8'*)

Return the assembled SBGN model as an XML string.

**Parameters** `pretty` (*Optional[bool]*) – If True, the SBGN string is formatted with indentation (for human viewing) otherwise no indentation is used. Default: True

**Returns** `sbgn_str` – An XML string representation of the SBGN model.

**Return type** `bytes` (str in Python 2)

**save\_model** (*file\_name='model.sbgn'*)

Save the assembled SBGN model in a file.

**Parameters** `file_name` (*Optional[str]*) – The name of the file to save the SBGN network to. Default: `model.sbgn`

### 4.8.8 Cytoscape JS networks (`indra.assemblers.cyjs.assembler`)

**class** `indra.assemblers.cyjs.assembler.CyJSAssembler` (*stmts=None*)

This class assembles a CytoscapeJS graph from a set of INDRA Statements.

CytoscapeJS is a web-based network library for analysis and visualisation: <http://js.cytoscape.org/>

**Parameters** `statements` (*Optional[list[indra.statements.Statement]]*) – A list of INDRA Statements to be assembled.

**statements**

A list of INDRA Statements to be assembled.

**Type** `list[indra.statements.Statement]`

**add\_statements** (*stmts*)

Add INDRA Statements to the assembler's list of statements.

**Parameters** `stmts` (*list[indra.statements.Statement]*) – A list of `indra.statements.Statement` to be added to the statement list of the assembler.

**get\_gene\_names** ()

Gather gene names of all nodes and node members

**make\_model** (*\*args, \*\*kwargs*)

Assemble a Cytoscape JS network from INDRA Statements.

This method assembles a Cytoscape JS network from the set of INDRA Statements added to the assembler.

**Parameters** `grouping` (*bool*) – If True, the nodes with identical incoming and outgoing edges are grouped and the corresponding edges are merged.

**Returns** `cyjs_str` – The json serialized Cytoscape JS model.

**Return type** `str`

**print\_cyjs\_context** ()

Return a list of node names and their respective context.

**Returns** `cyjs_str_context` – A json string of the context dictionary. e.g. - `{'CCLE' : {'bin_expression' : {'cell_line1' : {'gene1':'val1'} }, 'bin_expression' : {'cell_line' : {'gene1':'val1'} }}}`

**Return type** `str`

**print\_cyjs\_graph** ()

Return the assembled Cytoscape JS network as a json string.

**Returns** `cyjs_str` – A json string representation of the Cytoscape JS network.

**Return type** `str`

**save\_json** (*fname\_prefix='model'*)

Save the assembled Cytoscape JS network in a json file.

This method saves two files based on the file name prefix given. It saves one json file with the graph itself, and another json file with the context.

**Parameters** `fname_prefix` (*Optional[str]*) – The prefix of the files to save the Cytoscape JS network and context to. Default: `model`

**save\_model** (*fname='model.js'*)

Save the assembled Cytoscape JS network in a js file.

**Parameters** `file_name` (*Optional[str]*) – The name of the file to save the Cytoscape JS network to. Default: model.js

**set\_CCLE\_context** (*cell\_types*)  
Set context of all nodes and node members from CCLE.

#### 4.8.9 Causal analysis graphs (`indra.assemblers.cag.assembler`)

**class** `indra.assemblers.cag.assembler.CAGAssembler` (*stmts=None*)  
Assembles a causal analysis graph from INDRA Statements.

**Parameters** `stmts` (*Optional[list[indra.statement.Statements]]*) – A list of INDRA Statements to be assembled. Currently supports Influence Statements.

**statements**  
A list of INDRA Statements to be assembled.

**Type** `list[indra.statements.Statement]`

**CAG**  
A networkx MultiDiGraph object representing the causal analysis graph.

**Type** `nx.MultiDiGraph`

**add\_statements** (*stmts*)  
Add a list of Statements to the assembler.

**export\_to\_cytoscapejs** ()  
Return CAG in format readable by CytoscapeJS.

**Returns** A JSON-like dict representing the graph for use with CytoscapeJS.

**Return type** `dict`

**generate\_jupyter\_js** (*cyjs\_style=None, cyjs\_layout=None*)  
Generate Javascript from a template to run in Jupyter notebooks.

**Parameters**

- **cyjs\_style** (*Optional[dict]*) – A dict that sets CytoscapeJS style as specified in <https://github.com/cytoscape/cytoscape.js/blob/master/documentation/md/style.md>.
- **cyjs\_layout** (*Optional[dict]*) – A dict that sets CytoscapeJS layout parameters.

**Returns** A Javascript string to be rendered in a Jupyter notebook cell.

**Return type** `str`

**make\_model** (*grounding\_ontology='UN', grounding\_threshold=None*)  
Return a networkx MultiDiGraph representing a causal analysis graph.

**Parameters**

- **grounding\_ontology** (*Optional[str]*) – The ontology from which the grounding should be taken (e.g. UN, FAO)
- **grounding\_threshold** (*Optional[float]*) – Minimum threshold score for Eidos grounding.

**Returns** The assembled CAG.

**Return type** `nx.MultiDiGraph`

#### 4.8.10 Tabular output (`indra.assemblers.tsv.assembler`)

**class** `indra.assemblers.tsv.assembler.TsvAssembler` (*statements=None*)

Assembles Statements into a set of tabular files for export or curation.

Currently designed for use with “raw” Statements, i.e., Statements with a single evidence entry. Exports Statements into a single tab-separated file with the following columns:

**INDEX** A 1-indexed integer identifying the statement.

**UUID** The UUID of the Statement.

**TYPE** Statement type, given by the name of the class in `indra.statements`.

**STR** String representation of the Statement. Contains most relevant information for curation including any additional statement data beyond the Statement type and Agents.

**AG\_A\_TEXT** For Statements extracted from text, the text in the sentence corresponding to the first agent (i.e., the ‘TEXT’ entry in the `db_refs` dictionary). For all other Statements, the Agent name is given. Empty field if the Agent is None.

**AG\_A\_LINKS** Groundings for the first agent given as a comma-separated list of `identifiers.org` links. Empty if the Agent is None.

**AG\_A\_STR** String representation of the first agent, including additional agent context (e.g. modification, mutation, location, and bound conditions). Empty if the Agent is None.

**AG\_B\_TEXT, AG\_B\_LINKS, AG\_B\_STR** As above for the second agent. Note that the Agent may be None (and these fields left empty) if the Statement consists only of a single Agent (e.g., SelfModification, ActiveForm, or Translocation statement).

**PMID** PMID of the first entry in the evidence list for the Statement.

**TEXT** Evidence text for the Statement.

**IS\_HYP** Whether the Statement represents a “hypothesis”, as flagged by some reading systems and recorded in the `evidence.epistemics[‘hypothesis’]` field.

**IS\_DIRECT** Whether the Statement represents a direct physical interactions, as recorded by the `evidence.epistemics[‘direct’]` field.

In addition, if the `add_curation_cols` flag is set when calling `TsvAssembler.make_model()`, the following additional (empty) columns will be added, to be filled out by curators:

**AG\_A\_IDS\_CORRECT** Correctness of Agent A grounding.

**AG\_A\_STATE\_CORRECT** Correctness of Agent A context (e.g., modification, bound, and other conditions).

**AG\_B\_IDS\_CORRECT, AG\_B\_STATE\_CORRECT** As above, for Agent B.

**EVENT\_CORRECT** Whether the event is supported by the evidence text if the entities (Agents A and B) are considered as placeholders (i.e., ignoring the correctness of their grounding).

**RES\_CORRECT** For Modification statements, whether the amino acid residue indicated by the Statement is supported by the evidence.

**POS\_CORRECT** For Modification statements, whether the amino acid position indicated by the Statement is supported by the evidence.

**SUBJ\_ACT\_CORRECT** For Activation/Inhibition Statements, whether the activity indicated for the subject (Agent A) is supported by the evidence.

**OBJ\_ACT\_CORRECT** For Activation/Inhibition Statements, whether the activity indicated for the object (Agent B) is supported by the evidence.

**HYP\_CORRECT** Whether the Statement is correctly flagged as a hypothesis.

**HYP\_CORRECT** Whether the Statement is correctly flagged as direct.

**Parameters** **stmts** (*Optional[list[indra.statements.Statement]]*) – A list of INDRA Statements to be assembled.

**statements**

A list of INDRA Statements to be assembled.

**Type** list[indra.statements.Statement]

**make\_model** (*output\_file, add\_curation\_cols=False, up\_only=False*)

Export the statements into a tab-separated text file.

**Parameters**

- **output\_file** (*str*) – Name of the output file.
- **add\_curation\_cols** (*bool*) – Whether to add columns to facilitate statement curation. Default is False (no additional columns).
- **up\_only** (*bool*) – Whether to include identifiers.org links *only* for the Uniprot grounding of an agent when one is available. Because most spreadsheets allow only a single hyperlink per cell, this can make it easier to link to Uniprot information pages for curation purposes. Default is False.

#### 4.8.11 HTML browsing and curation (`indra.assemblers.html.assembler`)

Format a set of INDRA Statements into an HTML-formatted report which also supports curation.

```
class indra.assemblers.html.assembler.HtmlAssembler (statements=None, summary_metadata=None, ev_totals=None, source_counts=None, title='INDRA Results', db_rest_url=None)
```

Generates an HTML-formatted report from INDRA Statements.

The HTML report format includes statements formatted in English (by the EnglishAssembler), text and metadata for the Evidence object associated with each Statement, and a Javascript-based curation interface linked to the INDRA database (access permitting). The interface allows for curation of statements at the evidence level by letting the user specify type of error and (optionally) provide a short description of the error.

**Parameters**

- **statements** (*Optional[list[indra.statements.Statement]]*) – A list of INDRA Statements to be added to the assembler. Statements can also be added using the `add_statements` method after the assembler has been instantiated.
- **summary\_metadata** (*Optional[dict]*) – Dictionary of statement corpus metadata such as that provided by the INDRA REST API. Default is None. Each value should be a concise summary of O(1), not of order the length of the list, such as the evidence totals. The keys should be informative human-readable strings.
- **ev\_totals** (*Optional[dict]*) – A dictionary of the total evidence available for each statement indexed by hash. Default: None
- **source\_counts** (*Optional[dict]*) – A dictionary of the itemized evidence counts, by source, available for each statement, indexed by hash. Default: None.

- **title** (*str*) – The title to be printed at the top of the page.
- **db\_rest\_url** (*Optional[str]*) – The URL to a DB REST API to use for links out to further evidence. If given, this URL will be prepended to links that load additional evidence for a given Statement. One way to obtain this value is from the configuration entry `indra.config.get_config('INDRA_DB_REST_URL')`. If None, the URLs are constructed as relative links. Default: None

**statements**

A list of INDRA Statements to assemble.

**Type** `list[indra.statements.Statement]`

**model**

The HTML report formatted as a single string.

**Type** `str`

**metadata**

Dictionary of statement list metadata such as that provided by the INDRA REST API.

**Type** `dict`

**ev\_totals**

A dictionary of the total evidence available for each statement indexed by hash.

**Type** `dict`

**db\_rest\_url**

The URL to a DB REST API.

**Type** `str`

**add\_statements** (*statements*)

Add a list of Statements to the assembler.

**Parameters** **statements** (*list[indra.statements.Statement]*) – A list of INDRA Statements to be added to the assembler.

**append\_warning** (*msg*)

Append a warning message to the model to expose issues.

**make\_model** (*template=None, \*\*template\_kwargs*)

Return the assembled HTML content as a string.

**Returns** The assembled HTML as a string.

**Return type** `str`

**save\_model** (*fname*)

Save the assembled HTML into a file.

**Parameters** **fname** (*str*) – The path to the file to save the HTML into.

**class** `indra.assemblers.html.assembler.IndraHTMLLoader` (*root\_paths=None*)

A home-grown template loader to load the INDRA templates.

Based on the example found here: <http://jinja.pocoo.org/docs/2.10/api/#loaders>

**Parameters** **root\_paths** (*dict*) – A dict of strings indicating possible roots for template directories, keyed by basename. To be more specific, an entry `{'indra': '/path/to/module'}` would mean that you expect to have `'indra/template.html'` mapped to `'/path/to/module/templates/template.html'`. The default is `{'indra': HERE}`.

**get\_source** (*environment, template*)

Get the template source, filename and reload helper for a template. It's passed the environment and template name and has to return a tuple in the form (source, filename, uptodate) or raise a *TemplateNotFound* error if it can't locate the template.

The source part of the returned tuple must be the source of the template as unicode string or a ASCII bytestring. The filename should be the name of the file on the filesystem if it was loaded from there, otherwise *None*. The filename is used by python for the tracebacks if no loader extension is used.

The last item in the tuple is the *uptodate* function. If auto reloading is enabled it's always called to check if the template changed. No arguments are passed so the function must store the old state somewhere (for example in a closure). If it returns *False* the template will be reloaded.

```
indra.assemblers.html.assembler.tag_text(text, tag_info_list)
```

Apply start/end tags to spans of the given text.

**Parameters**

- **text** (*str*) – Text to be tagged
- **tag\_info\_list** (*list of tuples*) – Each tuple refers to a span of the given text. Fields are (*start\_ix, end\_ix, substring, start\_tag, close\_tag*), where *substring*, *start\_tag*, and *close\_tag* are strings. If any of the given spans of text overlap, the longest span is used.

**Returns** String where the specified substrings have been surrounded by the given start and close tags.

**Return type** str

#### 4.8.12 BMI wrapper for PySB-assembled models (`indra.assemblers.pysb.bmi_wrapper`)

This module allows creating a Basic Modeling Interface (BMI) model from and automatically assembled PySB model. The BMI model can be instantiated within a simulation workflow system where it is simulated together with other models.

```
class indra.assemblers.pysb.bmi_wrapper.BMIModel(model, inputs=None,
                                                stop_time=1000, out-
                                                side_name_map=None)
```

This class represents a BMI model wrapping a model assembled by INDRA.

**Parameters**

- **model** (*pysb.Model*) – A PySB model assembled by INDRA to be wrapped in BMI.
- **inputs** (*Optional[list[str]]*) – A list of variable names that are considered to be inputs to the model meaning that they are read from other models. Note that designating a variable as input means that it must be provided by another component during the simulation.
- **stop\_time** (*int*) – The stopping time for this model, controlling the time units up to which the model is simulated.
- **outside\_name\_map** (*dict*) – A dictionary mapping outside variables names to inside variable names (i.e. ones that are in the wrapped model)

**export\_into\_python** ()

Write the model into a pickle and create a module that loads it.

The model basically exports itself as a pickle file and a Python file is then written which loads the pickle file. This allows importing the model in the simulation workflow.

**finalize()**

Finish the simulation and clean up resources as needed.

**get\_attribute(*att\_name*)**

Return the value of a given attribute.

Attributes include: `model_name`, `version`, `author_name`, `grid_type`, `time_step_type`, `step_method`, `time_units`

**Parameters** `att_name` (*str*) – The name of the attribute whose value should be returned.

**Returns** `value` – The value of the attribute

**Return type** `str`

**get\_current\_time()**

Return the current time point that the model is at during simulation

**Returns** `time` – The current time point

**Return type** `float`

**get\_input\_var\_names()**

Return a list of variables names that can be set as input.

**Returns** `var_names` – A list of variable names that can be set from the outside

**Return type** `list[str]`

**get\_output\_var\_names()**

Return a list of variables names that can be read as output.

**Returns** `var_names` – A list of variable names that can be read from the outside

**Return type** `list[str]`

**get\_start\_time()**

Return the initial time point of the model.

**Returns** `start_time` – The initial time point of the model.

**Return type** `float`

**get\_status()**

Return the current status of the model.

**get\_time\_step()**

Return the time step associated with model simulation.

**Returns** `dt` – The time step for model simulation

**Return type** `float`

**get\_time\_units()**

Return the time units of the model simulation.

**Returns** `units` – The time unit of simulation as a string

**Return type** `str`

**get\_value(*var\_name*)**

Return the value of a given variable.

**Parameters** `var_name` (*str*) – The name of the variable whose value should be returned

**Returns** `value` – The value of the given variable in the current state

**Return type** `float`



**get\_values** (*var\_name*)

Return the value of a given variable.

**Parameters** **var\_name** (*str*) – The name of the variable whose value should be returned

**Returns** **value** – The value of the given variable in the current state

**Return type** float

**get\_var\_name** (*var\_name*)

Return the internal variable name given an outside variable name.

**Parameters** **var\_name** (*str*) – The name of the outside variable to map

**Returns** **internal\_var\_name** – The internal name of the corresponding variable

**Return type** str

**get\_var\_rank** (*var\_name*)

Return the matrix rank of the given variable.

**Parameters** **var\_name** (*str*) – The name of the variable whose rank should be returned

**Returns** **rank** – The dimensionality of the variable, 0 for scalar, 1 for vector, etc.

**Return type** int

**get\_var\_type** (*var\_name*)

Return the type of a given variable.

**Parameters** **var\_name** (*str*) – The name of the variable whose type should be returned

**Returns** **unit** – The type of the variable as a string

**Return type** str

**get\_var\_units** (*var\_name*)

Return the units of a given variable.

**Parameters** **var\_name** (*str*) – The name of the variable whose units should be returned

**Returns** **unit** – The units of the variable

**Return type** str

**initialize** (*cfg\_file=None, mode=None*)

Initialize the model for simulation, possibly given a config file.

**Parameters** **cfg\_file** (*Optional[str]*) – The name of the configuration file to load, optional.

**make\_repository\_component** ()

Return an XML string representing this BMI in a workflow.

This description is required by EMELI to discover and load models.

**Returns** **xml** – String serialized XML representation of the component in the model repository.

**Return type** str

**set\_value** (*var\_name, value*)

Set the value of a given variable to a given value.

**Parameters**

- **var\_name** (*str*) – The name of the variable in the model whose value should be set.
- **value** (*float*) – The value the variable should be set to

**set\_values** (*var\_name*, *value*)

Set the value of a given variable to a given value.

**Parameters**

- **var\_name** (*str*) – The name of the variable in the model whose value should be set.
- **value** (*float*) – The value the variable should be set to

**update** (*dt=None*)

Simulate the model for a given time interval.

**Parameters** **dt** (*Optional[float]*) – The time step to simulate, if None, the default built-in time step is used.

### 4.8.13 PyBEL graphs (`indra.assemblers.pybel.assembler`)

```
class indra.assemblers.pybel.assembler.PybelAssembler (stmts=None, name=None,
description=None, version=None, authors=None,
contact=None, license=None, copyright=None, disclaimer=None)
```

Assembles a PyBEL graph from a set of INDRA Statements.

PyBEL tools can subsequently be used to export the PyBEL graph into BEL script files, SIF files, and other related output formats.

**Parameters**

- **stmts** (`list[indra.statement.Statement]`) – The list of Statements to assemble.
- **name** (*str*) – Name of the assembled PyBEL network.
- **description** (*str*) – Description of the assembled PyBEL network.
- **version** (*str*) – Version of the assembled PyBEL network.
- **authors** (*str*) – Author(s) of the network.
- **contact** (*str*) – Contact information (email) of the responsible author.
- **license** (*str*) – License information for the network.
- **copyright** (*str*) – Copyright information for the network.
- **disclaimer** (*str*) – Any disclaimers for the network.

#### Examples

```
>>> from indra.statements import *
>>> map2k1 = Agent('MAP2K1', db_refs={'HGNC': '6840'})
>>> mapk1 = Agent('MAPK1', db_refs={'HGNC': '6871'})
>>> stmt = Phosphorylation(map2k1, mapk1, 'T', '185')
>>> pba = PybelAssembler([stmt])
>>> belgraph = pba.make_model()
>>> sorted(node.as_bel() for node in belgraph) # doctest:+IGNORE_UNICODE
['p(HGNC:MAP2K1)', 'p(HGNC:MAPK1)', 'p(HGNC:MAPK1, pmod(Ph, Thr, 185))']
>>> len(belgraph)
3
```

(continues on next page)

(continued from previous page)

```
>>> belgraph.number_of_edges()
2
```

**save\_model** (*path, output\_format=None*)

Save the `pybel.BELGraph` using one of the outputs from `pybel`

**Parameters**

- **path** (*str*) – The path to output to
- **output\_format** (*Optional[str]*) – Output format as `cx`, `pickle`, `json` or defaults to `bel`

**to\_database** (*manager=None*)

Send the model to the PyBEL database

This function wraps `pybel.to_database()`.

**Parameters** **manager** (*Optional[pybel.manager.Manager]*) – A PyBEL database manager. If none, first checks the PyBEL configuration for `PYBEL_CONNECTION` then checks the environment variable `PYBEL_REMOTE_HOST`. Finally, defaults to using SQLite database in PyBEL data directory (automatically configured by PyBEL)

**Returns** **network** – The SQLAlchemy model representing the network that was uploaded. Returns None if upload fails.

**Return type** `Optional[pybel.manager.models.Network]`

**to\_web** (*host=None, user=None, password=None*)

Send the model to BEL Commons by wrapping `pybel.to_web()`

The parameters `host`, `user`, and `password` all check the PyBEL configuration, which is located at `~/config/pybel/config.json` by default

**Parameters**

- **host** (*Optional[str]*) – The host name to use. If none, first checks the PyBEL configuration entry `PYBEL_REMOTE_HOST`, then the environment variable `PYBEL_REMOTE_HOST`. Finally, defaults to <https://bel-commons.scai.fraunhofer.de>.
- **user** (*Optional[str]*) – The username (email) to use. If none, first checks the PyBEL configuration entry `PYBEL_REMOTE_USER`, then the environment variable `PYBEL_REMOTE_USER`.
- **password** (*Optional[str]*) – The password to use. If none, first checks the PyBEL configuration entry `PYBEL_REMOTE_PASSWORD`, then the environment variable `PYBEL_REMOTE_PASSWORD`.

**Returns** **response** – The response from the BEL Commons network upload endpoint.

**Return type** `requests.Response`

`indra.assemblers.pybel.assembler.get_causal_edge(stmt, activates)`

Returns the causal, polar edge with the correct “contact”.

#### 4.8.14 Kami models (`indra.assemblers.kami.assembler`)

**class** `indra.assemblers.kami.assembler.KamiAssembler` (*statements=None*)

**make\_model** (*policies=None, initial\_conditions=True, reverse\_effects=False*)

Assemble the Kami model from the collected INDRA Statements.

This method assembles a Kami model from the set of INDRA Statements. The assembled model is both returned and set as the assembler's model argument.

#### Parameters

- **policies** (*Optional[Union[str, dict]]*) – A string or dictionary of policies, as defined in `indra.assemblers.KamiAssembler`. This set of policies locally supersedes the default setting in the assembler. This is useful when this function is called multiple times with different policies.
- **initial\_conditions** (*Optional[bool]*) – If True, default initial conditions are generated for the agents in the model.

**Returns model** – The assembled Kami model.

**Return type** dict

**class** `indra.assemblers.kami.assembler.Nugget` (*id, name, rate*)

Represents a Kami Nugget.

**add\_agent** (*agent*)

Add an INDRA Agent and its conditions to the Nugget.

**add\_edge** (*from\_node, to\_node*)

Add an edge between two nodes to the Nugget.

**add\_node** (*name\_base, attrs=None*)

Add a node with a given base name to the Nugget and return ID.

**add\_typing** (*node\_id, typing*)

Add typing information to a node in the Nugget.

**get\_nugget\_dict** ()

Return the Nugget as a dictionary.

**get\_typing\_dict** ()

Return the Nugget's typing information as a dictionary.

### 4.8.15 IndraNet Graphs (`indra.assemblers.indranet`)

A Networkx Representation of INDRA Statements

**class** `indra.assemblers.indranet.net.IndraNet` (*incoming\_graph\_data=None, \*\*attr*)

A Networkx representation of INDRA Statements.

**classmethod digraph\_from\_df** (*df, flattening\_method=None, weight\_mapping=None*)

Create a digraph from a pandas DataFrame.

**classmethod from\_df** (*df*)

Create an IndraNet MultiDiGraph from a pandas DataFrame.

Returns an instance of IndraNet with graph data filled out from a dataframe containing pairwise interactions.

**Parameters df** (*pd.DataFrame*) – A `pandas.DataFrame` with each row containing node and edge data for one edge. Indices are used to distinguish multiedges between a pair of nodes. Any columns not part of the mandatory columns are considered extra attributes. Columns starting with 'agA\_' or 'agB\_' (excluding the agA/B\_name) will be added to its respective nodes as node attributes. Any other columns will be added as edge attributes.

**Returns** An IndraNet object

**Return type** *IndraNet*

**classmethod** `signed_from_df` (*df*, *sign\_dict=None*, *flattening\_method=None*,  
*weight\_mapping=None*)

Create a signed graph from a pandas DataFrame.

**to\_digraph** (*flattening\_method=None*, *weight\_mapping=None*)

Flatten the IndraNet to a DiGraph

**Parameters**

- **flattening\_method** (*str/function*) – The method to use when updating the belief for the flattened edge
- **weight\_mapping** (*function*) – A function taking at least the graph G as an argument and returning G after adding edge weights as an edge attribute to the flattened edges using the reserved keyword ‘weight’.

**Returns** G – An IndraNet graph flattened to a DiGraph

**Return type** *IndraNet*(*nx.DiGraph*)

**to\_signed\_graph** (*sign\_dict=None*, *flattening\_method=None*, *weight\_mapping=None*)

Flatten the IndraNet to a signed graph.

**Parameters**

- **sign\_dict** (*dict*) – A dictionary mapping a Statement type to a sign to be used for the edge. By default only Activation and IncreaseAmount are added as positive edges and Inhibition and DecreaseAmount are added as negative edges, but a user can pass any other Statement types in a dictionary.
- **flattening\_method** (*str/function(G, edge)*) – The method to use when updating the belief for the flattened edge.

If a string is provided, it must be one of the predefined options ‘simple\_scorer’ or ‘complementary\_belief’.

If a function is provided, it must take the flattened graph ‘G’ and an edge ‘edge’ to perform the belief flattening on and return a number:

```
>>> def flattening_function(G, edge):
...     # Return the average belief score of the constituent edges
...     all_beliefs = [s['belief']
...                     for s in G.edges[edge]['statements']]
...     return sum(all_beliefs)/len(all_beliefs)
```

- **weight\_mapping** (*function(G)*) – A function taking at least the graph G as an argument and returning G after adding edge weights as an edge attribute to the flattened edges using the reserved keyword ‘weight’.

Example:

```
>>> def weight_mapping(G):
...     # Sets the flattened weight to the average of the
...     # inverse source count
...     for edge in G.edges:
...         w = [1/s['evidence_count']
...               for s in G.edges[edge]['statements']]
...         G.edges[edge]['weight'] = sum(w)/len(w)
...     return G
```

**Returns** `SG` – An IndraNet graph flattened to a signed graph

**Return type** `IndraNet`(`nx.MultiDiGraph`)

**class** `indra.assemblers.indranet.assembler.IndraNetAssembler` (`statements=None`)  
Assembler to create an IndraNet object from a list of INDRA statements.

**Parameters** `statements` (`list[indra.statements.Statement]`) – A list of INDRA Statements to be assembled.

**model**

An IndraNet graph object assembled by this class.

**Type** `IndraNet`

**add\_statements** (`stmts`)

Add INDRA Statements to the assembler’s list of statements.

**Parameters** `stmts` (`list[indra.statements.Statement]`) – A list of `indra.statements.Statement` to be added to the statement list of the assembler.

**make\_df** (`exclude_stmts=None`, `complex_members=3`)

Create a dataframe containing information extracted from assembler’s list of statements necessary to build an IndraNet.

**Parameters**

- **exclude\_stmts** (`list[str]`) – A list of statement type names to not include into a dataframe.
- **complex\_members** (`int`) – Maximum allowed size of a complex to be included in the data frame. All complexes larger than `complex_members` will be rejected. For accepted complexes, all permutations of their members will be added as dataframe records.

**Returns** `df` – Pandas DataFrame object containing information extracted from statements.

**Return type** `pd.DataFrame`

**make\_model** (`exclude_stmts=None`, `complex_members=3`, `graph_type='multi_graph'`,  
`sign_dict=None`, `belief_flattening=None`, `weight_flattening=None`)

Assemble an IndraNet graph object.

**Parameters**

- **exclude\_stmts** (`list[str]`) – A list of statement type names to not include in the graph.
- **complex\_members** (`int`) – Maximum allowed size of a complex to be included in the graph. All complexes larger than `complex_members` will be rejected. For accepted complexes, all permutations of their members will be added as edges.
- **graph\_type** (`str`) – Specify the type of graph to assemble. Chose from ‘multi\_graph’ (default), ‘digraph’, ‘signed’.
- **sign\_dict** (`dict`) – A dictionary mapping a Statement type to a sign to be used for the edge. This parameter is only used with the ‘signed’ option. See `IndraNet.to_signed_graph` for more info.
- **belief\_flattening** (`str|function(G, edge)`) – The method to use when updating the belief for the flattened edge.

If a string is provided, it must be one of the predefined options ‘simple\_scorer’ or ‘complementary\_belief’.

If a function is provided, it must take the flattened graph ‘G’ and an edge ‘edge’ to perform the belief flattening on and return a number:

```
>>> def belief_flattening(G, edge):
...     # Return the average belief score of the constituent edges
...     all_beliefs = [s['belief']]
...     for s in G.edges[edge]['statements']:
...         return sum(all_beliefs)/len(all_beliefs)
```

- **weight\_flattening** (*function(G)*) – A function taking at least the graph G as an argument and returning G after adding edge weights as an edge attribute to the flattened edges using the reserved keyword ‘weight’.

Example:

```
>>> def weight_flattening(G):
...     # Sets the flattened weight to the average of the
...     # inverse source count
...     for edge in G.edges:
...         w = [1/s['evidence_count']]
...         for s in G.edges[edge]['statements']:
...             G.edges[edge]['weight'] = sum(w)/len(w)
...     return G
```

Returns **model** – IndraNet graph object.

Return type *IndraNet*

`indra.assemblers.indranet.assembler.get_ag_ns_id(ag)`

Return a tuple of name space, id from an Agent’s db\_refs.

## 4.9 Explanation (`indra.explanation`)

### 4.9.1 Check whether a model satisfies a property (`indra.explanation.model_checker`)

Shared Model Checking Functionality (`indra.explanation.model_checker.model_checker`)

```
class indra.explanation.model_checker.model_checker.ModelChecker (model, state-
                                                                statements=None,
                                                                do_sampling=False,
                                                                seed=None)
```

The parent class of all ModelCheckers.

#### Parameters

- **model** (*pysb.Model or indra.assemblers.indranet.IndraNet or PyBEL.Model*) – Depending on the ModelChecker class, can be different type.
- **statements** (*Optional[list[indra.statements.Statement]]*) – A list of INDRA Statements to check the model against.
- **do\_sampling** (*bool*) – Whether to use breadth-first search or weighted sampling to generate paths. Default is False (breadth-first search).
- **seed** (*int*) – Random seed for sampling (optional, default is None).

**graph**

A DiGraph with signed nodes to find paths in.

**Type** `nx.DiGraph`

**add\_statements** (*stmts*)

Add to the list of statements to check against the model.

**Parameters** *stmts* (`list[indra.statements.Statement]`) – The list of Statements to be added for checking.

**check\_model** (*max\_paths=1, max\_path\_length=5*)

Check all the statements added to the ModelChecker.

**Parameters**

- **max\_paths** (`Optional[int]`) – The maximum number of specific paths to return for each Statement to be explained. Default: 1
- **max\_path\_length** (`Optional[int]`) – The maximum length of specific paths to return. Default: 5

**Returns** Each tuple contains the Statement checked against the model and a PathResult object describing the results of model checking.

**Return type** list of (`Statement`, `PathResult`)

**check\_statement** (*stmt, max\_paths=1, max\_path\_length=5*)

Check a single Statement against the model.

**Parameters**

- **stmt** (`indra.statements.Statement`) – The Statement to check.
- **max\_paths** (`Optional[int]`) – The maximum number of specific paths to return for each Statement to be explained. Default: 1
- **max\_path\_length** (`Optional[int]`) – The maximum length of specific paths to return. Default: 5

**Returns** **result** – A PathResult object containing the result of a test.

**Return type** `indra.explanation.modelchecker.PathResult`

**find\_paths** (*subj, obj, max\_paths=1, max\_path\_length=5*)

Check for a source/target path in the model.

**Parameters**

- **subj** (`pysb.MonomerPattern` or `tuple`) – Relevant to the model information about the subject of the Statement being checked (monomer pattern in PySB, source node for other models).
- **obj** (`tuple`) – Tuple representing the target node (created from PySB model Observable, PyBEL node, or Agent.name with a target sign).
- **max\_paths** (`int`) – The maximum number of specific paths to return.
- **max\_path\_length** (`int`) – The maximum length of specific paths to return.

**Returns** PathResult object indicating the results of the attempt to find a path.

**Return type** `PathResult`

**get\_graph** (*\*\*kwargs*)

Return a graph with signed nodes to find the path.



**process\_statement** (*stmt*)

This method processes the test statement to get the data about subject and object, according to the specific model requirements for model checking, e.g. PysbModelChecker gets subject monomer patterns and observables, while graph based ModelCheckers will return signed nodes corresponding to subject and object. If any of the requirements are not satisfied, result code is also returned to construct PathResult object.

**Parameters** *stmt* (*indra.statements.Statement*) – A statement to process.

**Returns**

- **subj\_data** (*list or None*) – Data about statement subject to be used as source nodes.
- **obj\_data** (*list or None*) – Data about statement object to be used as target nodes.
- **result\_code** (*str or None*) – Result code to construct PathResult.

**process\_subject** (*subj\_data*)

Processes the subject of the test statement and returns the necessary information to check the statement. In case of PysbModelChecker, method returns *input\_rule\_set*. If any of the requirements are not satisfied, result code is also returned to construct PathResult object.

**prune\_nodes** (*graph*)

Prune nodes with sign (1) if they do not have predecessors.

**signed\_edges\_to\_signed\_nodes** (*graph, prune\_nodes=True, edge\_signs={'neg': 1, 'pos': 0}*)

Convert a graph with signed edges to a graph with signed nodes.

Each pair of nodes linked by an edge in an input graph are represented as four nodes and two edges in the new graph. For example, an edge (a, b, 0), where a and b are nodes and 0 is a sign of an edge (positive), will be represented as edges ((a, 0), (b, 0)) and ((a, 1), (b, 1)), where (a, 0), (a, 1), (b, 0), (b, 1) are signed nodes. An edge (a, b, 1) with sign 1 (negative) will be represented as edges ((a, 0), (b, 1)) and ((a, 1), (b, 0)).

**Parameters**

- **graph** (*networkx.MultiDiGraph*) – Graph with signed edges to convert. Can have multiple edges between a pair of nodes.
- **prune\_nodes** (*Optional[bool]*) – If True, iteratively prunes negative (with sign 1) nodes without predecessors.
- **edge\_signs** (*dict*) – A dictionary representing the signing policy of incoming graph. The dictionary should have strings 'pos' and 'neg' as keys and integers as values.

**Returns signed\_nodes\_graph**

**Return type** *networkx.DiGraph*

```
class indra.explanation.model_checker.model_checker.PathMetric (source_node,
                                                    target_node,
                                                    length)
```

Describes results of simple path search (path existence).

**source\_node**

The source node of the path

**Type** *str*

**target\_node**

The target node of the path

**Type** *str*

**length**

The length of the path

**Type** int

**class** `indra.explanation.model_checker.model_checker.PathResult` (*path\_found*,  
*result\_code*,  
*max\_paths*,  
*max\_path\_length*)

Describes results of running the ModelChecker on a single Statement.

**path\_found**

True if a path was found, False otherwise.

**Type** bool

**result\_code**

- *STATEMENT\_TYPE\_NOT\_HANDLED* - The provided statement type is not handled
- *SUBJECT\_MONOMERS\_NOT\_FOUND* or *SUBJECT\_NOT\_FOUND* - Statement subject not found in model
- *OBSERVABLES\_NOT\_FOUND* or *OBJECT\_NOT\_FOUND* - Statement has no associated observable
- *NO\_PATHS\_FOUND* - Statement has no path for any observable
- *MAX\_PATH\_LENGTH\_EXCEEDED* - Statement has no path len <= MAX\_PATH\_LENGTH
- *PATHS\_FOUND* - Statement has path len <= MAX\_PATH\_LENGTH
- *INPUT\_RULES\_NOT\_FOUND* - No rules with Statement subject found
- *MAX\_PATHS\_ZERO* - Path found but MAX\_PATHS is set to zero

**Type** string

**max\_paths**

The maximum number of specific paths to return for each Statement to be explained.

**Type** int

**max\_path\_length**

The maximum length of specific paths to return.

**Type** int

**path\_metrics**

A list of PathMetric objects, each describing the results of a simple path search (path existence).

**Type** list[`indra.explanation.model_checker.PathMetric`]

**paths**

A list of paths obtained from path finding. Each path is a list of tuples (which are edges in the path), with the first element of the tuple the name of a rule, and the second element its polarity in the path.

**Type** list[list[tuple[str, int]]]

## Checking PySB model (`indra.explanation.model_checker.pysb`)

```
class indra.explanation.model_checker.pysb.PysbModelChecker (model, state-
                                                    ments=None,
                                                    agent_obs=None,
                                                    do_sampling=False,
                                                    seed=None,
                                                    model_stmts=None)
```

Check a PySB model against a set of INDRA statements.

### Parameters

- **model** (*pysb.Model*) – A PySB model to check.
- **statements** (*Optional[list[indra.statements.Statement]]*) – A list of INDRA Statements to check the model against.
- **agent\_obs** (*Optional[list[indra.statements.Agent]]*) – A list of INDRA Agents in a given state to be observed.
- **do\_sampling** (*bool*) – Whether to use breadth-first search or weighted sampling to generate paths. Default is False (breadth-first search).
- **seed** (*int*) – Random seed for sampling (optional, default is None).
- **model\_stmts** (*list[indra.statements.Statement]*) – A list of INDRA statements used to assemble PySB model.

**draw\_im** (*fname*)

Draw and save the influence map in a file.

**Parameters** **fname** (*str*) – The name of the file to save the influence map in. The extension of the file will determine the file format, typically png or pdf.

**generate\_im** (*model*)

Return a graph representing the influence map generated by Kappa

**Parameters** **model** (*pysb.Model*) – The PySB model whose influence map is to be generated

**Returns** **graph** – A MultiDiGraph representing the influence map

**Return type** `networkx.MultiDiGraph`

**get\_graph** (*prune\_im=True, prune\_im\_degrade=True, prune\_im\_subj\_obj=False*)

Get influence map and convert it to a graph with signed nodes.

**get\_im** (*force\_update=False*)

Get the influence map for the model, generating it if necessary.

**Parameters** **force\_update** (*bool*) – Whether to generate the influence map when the function is called. If False, returns the previously generated influence map if available. Defaults to True.

### Returns

The influence map can be rendered as a pdf using the dot layout program as follows:

```
im_agraph = nx.nx_agraph.to_agraph(influence_map)
im_agraph.draw('influence_map.pdf', prog='dot')
```

**Return type** `networkx MultiDiGraph` object containing the influence map.

**process\_statement** (*stmt*)

This method processes the test statement to get the data about subject and object, according to the specific model requirements for model checking, e.g. PysbModelChecker gets subject monomer patterns and observables, while graph based ModelCheckers will return signed nodes corresponding to subject and object. If any of the requirements are not satisfied, result code is also returned to construct PathResult object.

**Parameters** **stmt** (*indra.statements.Statement*) – A statement to process.

**Returns**

- **subj\_data** (*list or None*) – Data about statement subject to be used as source nodes.
- **obj\_data** (*list or None*) – Data about statement object to be used as target nodes.
- **result\_code** (*str or None*) – Result code to construct PathResult.

**process\_subject** (*subj\_mp*)

Processes the subject of the test statement and returns the necessary information to check the statement. In case of PysbModelChecker, method returns `input_rule_set`. If any of the requirements are not satisfied, result code is also returned to construct PathResult object.

**prune\_influence\_map** ()

Remove edges between rules causing problematic non-transitivity.

First, all self-loops are removed. After this initial step, edges are removed between rules when they share *all* child nodes except for each other; that is, they have a mutual relationship with each other and share all of the same children.

Note that edges must be removed in batch at the end to prevent edge removal from affecting the lists of rule children during the comparison process.

**prune\_influence\_map\_degrade\_bind\_positive** (*model\_stmts*)

Prune positive edges between X degrading and X forming a complex with Y.

**prune\_influence\_map\_subj\_obj** ()

Prune influence map to include only edges where the object of the upstream rule matches the subject of the downstream rule.

**score\_paths** (*paths, agents\_values, loss\_of\_function=False, sigma=0.15, include\_final\_node=False*)

Return scores associated with a given set of paths.

**Parameters**

- **paths** (*list[list[tuple[str, int]]]*) – A list of paths obtained from path finding. Each path is a list of tuples (which are edges in the path), with the first element of the tuple the name of a rule, and the second element its polarity in the path.
- **agents\_values** (*dict[indra.statements.Agent, float]*) – A dictionary of INDRA Agents and their corresponding measured value in a given experimental condition.
- **loss\_of\_function** (*Optional[boolean]*) – If True, flip the polarity of the path. For instance, if the effect of an inhibitory drug is explained, set this to True. Default: False
- **sigma** (*Optional[float]*) – The estimated standard deviation for the normally distributed measurement error in the observation model used to score paths with respect to data. Default: 0.15
- **include\_final\_node** (*Optional[boolean]*) – Determines whether the final node of the path is included in the score. Default: False

`indra.explanation.model_checker.pysb.remove_im_params` (*model, im*)

Remove parameter nodes from the influence map.

**Parameters**

- **model** (*pysb.core.Model*) – PySB model.
- **im** (*networkx.MultiDiGraph*) – Influence map.

**Returns** Influence map with the parameter nodes removed.

**Return type** *networkx.MultiDiGraph*

**Checking Signed Graph (*indra.explanation.model\_checker.signed\_graph*)**

```
class indra.explanation.model_checker.signed_graph.SignedGraphModelChecker (model,
                                                                    state-
                                                                    ments=None,
                                                                    do_sampling=False,
                                                                    seed=None)
```

Check an signed MultiDiGraph against a set of INDRA statements.

**Parameters**

- **model** (*networkx.MultiDiGraph*) – Signed MultiDiGraph to check.
- **statements** (*Optional[list[indra.statements.Statement]]*) – A list of INDRA Statements to check the model against.
- **do\_sampling** (*bool*) – Whether to use breadth-first search or weighted sampling to generate paths. Default is False (breadth-first search).
- **seed** (*int*) – Random seed for sampling (optional, default is None).

**get\_graph()**

Return a graph with signed nodes to find the path.

**process\_statement (stmt)**

This method processes the test statement to get the data about subject and object, according to the specific model requirements for model checking, e.g. *PysbModelChecker* gets subject monomer patterns and observables, while graph based *ModelCheckers* will return signed nodes corresponding to subject and object. If any of the requirements are not satisfied, result code is also returned to construct *PathResult* object.

**Parameters** **stmt** (*indra.statements.Statement*) – A statement to process.

**Returns**

- **subj\_data** (*list or None*) – Data about statement subject to be used as source nodes.
- **obj\_data** (*list or None*) – Data about statement object to be used as target nodes.
- **result\_code** (*str or None*) – Result code to construct *PathResult*.

**process\_subject (subj)**

Processes the subject of the test statement and returns the necessary information to check the statement. In case of *PysbModelChecker*, method returns *input\_rule\_set*. If any of the requirements are not satisfied, result code is also returned to construct *PathResult* object.

### Checking Unsigned Graph (`indra.explanation.model_checker.unsigned_graph`)

```
class indra.explanation.model_checker.unsigned_graph.UnsignedGraphModelChecker (model,  
state-  
ments=None,  
do_sampling=False,  
seed=None)
```

Check an unsigned DiGraph against a set of INDRA statements.

#### Parameters

- **model** (*networkx.DiGraph*) – Unsigned DiGraph to check.
- **statements** (*Optional[list[indra.statements.Statement]]*) – A list of INDRA Statements to check the model against.
- **do\_sampling** (*bool*) – Whether to use breadth-first search or weighted sampling to generate paths. Default is False (breadth-first search).
- **seed** (*int*) – Random seed for sampling (optional, default is None).

#### `get_graph()`

Return a graph with signed nodes to find the path.

#### `process_statement(stmt)`

This method processes the test statement to get the data about subject and object, according to the specific model requirements for model checking, e.g. PysbModelChecker gets subject monomer patterns and observables, while graph based ModelCheckers will return signed nodes corresponding to subject and object. If any of the requirements are not satisfied, result code is also returned to construct PathResult object.

**Parameters** **stmt** (*indra.statements.Statement*) – A statement to process.

#### Returns

- **subj\_data** (*list or None*) – Data about statement subject to be used as source nodes.
- **obj\_data** (*list or None*) – Data about statement object to be used as target nodes.
- **result\_code** (*str or None*) – Result code to construct PathResult.

#### `process_subject(subj)`

Processes the subject of the test statement and returns the necessary information to check the statement. In case of PysbModelChecker, method returns `input_rule_set`. If any of the requirements are not satisfied, result code is also returned to construct PathResult object.

### Checking PyBEL Graph (`indra.explanation.model_checker.pybel`)

```
class indra.explanation.model_checker.pybel.PybelModelChecker (model, state-  
ments=None,  
do_sampling=False,  
seed=None)
```

Check a PyBEL model against a set of INDRA statements.

#### Parameters

- **model** (*pybel.BELGraph*) – A Pybel model to check.
- **statements** (*Optional[list[indra.statements.Statement]]*) – A list of INDRA Statements to check the model against.
- **do\_sampling** (*bool*) – Whether to use breadth-first search or weighted sampling to generate paths. Default is False (breadth-first search).

- **seed** (*int*) – Random seed for sampling (optional, default is None).

#### **get\_graph()**

Convert a PyBELGraph to a graph with signed nodes.

#### **process\_statement** (*stmt*)

This method processes the test statement to get the data about subject and object, according to the specific model requirements for model checking, e.g. PysbModelChecker gets subject monomer patterns and observables, while graph based ModelCheckers will return signed nodes corresponding to subject and object. If any of the requirements are not satisfied, result code is also returned to construct PathResult object.

**Parameters** **stmt** (*indra.statements.Statement*) – A statement to process.

#### **Returns**

- **subj\_data** (*list or None*) – Data about statement subject to be used as source nodes.
- **obj\_data** (*list or None*) – Data about statement object to be used as target nodes.
- **result\_code** (*str or None*) – Result code to construct PathResult.

#### **process\_subject** (*subj*)

Processes the subject of the test statement and returns the necessary information to check the statement. In case of PysbModelChecker, method returns input\_rule\_set. If any of the requirements are not satisfied, result code is also returned to construct PathResult object.

## 4.10 Tools (`indra.tools`)

### 4.10.1 Run assembly components in a pipeline (`indra.tools.assemble_corpus`)

`indra.tools.assemble_corpus.align_statements` (*stmts1, stmts2, keyfun=None*)

Return alignment of two lists of statements by key.

#### **Parameters**

- **stmts1** (*list[indra.statements.Statement]*) – A list of INDRA Statements to align
- **stmts2** (*list[indra.statements.Statement]*) – A list of INDRA Statements to align
- **keyfun** (*Optional[function]*) – A function that takes a Statement as an argument and returns a key to align by. If not given, the default key function is a tuple of the names of the Agents in the Statement.

**Returns** **matches** – A list of tuples where each tuple has two elements, the first corresponding to an element of the `stmts1` list and the second corresponding to an element of the `stmts2` list. If a given element is not matched, its corresponding pair in the tuple is None.

**Return type** list(tuple)

`indra.tools.assemble_corpus.dump_statements` (*stmts, fname, protocol=4*)

Dump a list of statements into a pickle file.

#### **Parameters**

- **fname** (*str*) – The name of the pickle file to dump statements into.
- **protocol** (*Optional[int]*) – The pickle protocol to use (use 2 for Python 2 compatibility). Default: 4

`indra.tools.assemble_corpus.dump_stmt_strings(stmts, fname)`

Save printed statements in a file.

#### Parameters

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to save in a text file.
- **fname** (*Optional[str]*) – The name of a text file to save the printed statements into.

`indra.tools.assemble_corpus.expand_families(stmts_in, **kwargs)`

Expand FamPlex Agents to individual genes.

#### Parameters

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to expand.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.

**Returns** `stmts_out` – A list of expanded statements.

**Return type** `list[indra.statements.Statement]`

`indra.tools.assemble_corpus.filter_belief(stmts_in, belief_cutoff, **kwargs)`

Filter to statements with belief above a given cutoff.

#### Parameters

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **belief\_cutoff** (*float*) – Only statements with belief above the belief\_cutoff will be returned. Here  $0 < \text{belief\_cutoff} < 1$ .
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.

**Returns** `stmts_out` – A list of filtered statements.

**Return type** `list[indra.statements.Statement]`

`indra.tools.assemble_corpus.filter_by_db_refs(stmts_in, namespace, values, policy, **kwargs)`

Filter to Statements whose agents are grounded to a matching entry.

Statements are filtered so that the db\_refs entry (of the given namespace) of their Agent/Concept arguments take a value in the given list of values.

#### Parameters

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of Statements to filter.
- **namespace** (*str*) – The namespace in db\_refs to which the filter should apply.
- **values** (*list[str]*) – A list of values in the given namespace to which the filter should apply.
- **policy** (*str*) – The policy to apply when filtering for the db\_refs. “one”: keep Statements that contain at least one of the list of db\_refs and possibly others not in the list “all”: keep Statements that only contain db\_refs given in the list
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.
- **invert** (*Optional[bool]*) – If True, the Statements that do not match according to the policy are returned. Default: False



- **match\_suffix** (*Optional[bool]*) – If True, the suffix of the db\_refs entry is matches against the list of entries

**Returns** `stmts_out` – A list of filtered Statements.

**Return type** `list[indra.statements.Statement]`

`indra.tools.assemble_corpus.filter_by_type` (*stmts\_in, stmt\_type, \*\*kwargs*)

Filter to a given statement type.

#### Parameters

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **stmt\_type** (*indra.statements.Statement*) – The class of the statement type to filter for. Example: `indra.statements.Modification`
- **invert** (*Optional[bool]*) – If True, the statements that are not of the given type are returned. Default: False
- **save** (*Optional[str]*) – The name of a pickle file to save the results (`stmts_out`) into.

**Returns** `stmts_out` – A list of filtered statements.

**Return type** `list[indra.statements.Statement]`

`indra.tools.assemble_corpus.filter_concept_names` (*stmts\_in, name\_list, policy, \*\*kwargs*)

Return Statements that refer to concepts/agents given as a list of names.

#### Parameters

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of Statements to filter.
- **name\_list** (*list[str]*) – A list of concept/agent names to filter for.
- **policy** (*str*) – The policy to apply when filtering for the list of names. “one”: keep Statements that contain at least one of the list of names and possibly others not in the list “all”: keep Statements that only contain names given in the list
- **save** (*Optional[str]*) – The name of a pickle file to save the results (`stmts_out`) into.
- **invert** (*Optional[bool]*) – If True, the Statements that do not match according to the policy are returned. Default: False

**Returns** `stmts_out` – A list of filtered Statements.

**Return type** `list[indra.statements.Statement]`

`indra.tools.assemble_corpus.filter_direct` (*stmts\_in, \*\*kwargs*)

Filter to statements that are direct interactions

#### Parameters

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (`stmts_out`) into.

**Returns** `stmts_out` – A list of filtered statements.

**Return type** `list[indra.statements.Statement]`

`indra.tools.assemble_corpus.filter_enzyme_kinase` (*stmts\_in, \*\*kwargs*)

Filter Phosphorylations to ones where the enzyme is a known kinase.

**Parameters**

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.

**Returns** `stmts_out` – A list of filtered statements.

**Return type** `list[indra.statements.Statement]`

```
indra.tools.assemble_corpus.filter_evidence_source(stmts_in, source_apis, policy='one', **kwargs)
```

Filter to statements that have evidence from a given set of sources.

**Parameters**

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **source\_apis** (*list[str]*) – A list of sources to filter for. Examples: biopax, bel, reach
- **policy** (*Optional[str]*) – If 'one', a statement that has evidence from any of the sources is kept. If 'all', only those statements are kept which have evidence from all the input sources specified in source\_apis. If 'none', only those statements are kept that don't have evidence from any of the sources specified in source\_apis.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.

**Returns** `stmts_out` – A list of filtered statements.

**Return type** `list[indra.statements.Statement]`

```
indra.tools.assemble_corpus.filter_gene_list(stmts_in, gene_list, policy, allow_families=False, **kwargs)
```

Return statements that contain genes given in a list.

**Parameters**

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **gene\_list** (*list[str]*) – A list of gene symbols to filter for.
- **policy** (*str*) – The policy to apply when filtering for the list of genes. "one": keep statements that contain at least one of the list of genes and possibly others not in the list "all": keep statements that only contain genes given in the list
- **allow\_families** (*Optional[bool]*) – Will include statements involving FamPlex families containing one of the genes in the gene list. Default: False
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.
- **remove\_bound** (*Optional[str]*) – If true, removes bound conditions that are not genes in the list. If false (default), looks at agents in the bound conditions in addition to those participating in the statement directly when applying the specified policy.
- **invert** (*Optional[bool]*) – If True, the statements that do not match according to the policy are returned. Default: False

**Returns** `stmts_out` – A list of filtered statements.

**Return type** `list[indra.statements.Statement]`

`indra.tools.assemble_corpus.filter_genes_only(stmts_in, **kwargs)`

Filter to statements containing genes only.

#### Parameters

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **specific\_only** (*Optional[bool]*) – If True, only elementary genes/proteins will be kept and families will be filtered out. If False, families are also included in the output. Default: False
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.
- **remove\_bound** (*Optional[bool]*) – If true, removes bound conditions that are not genes. If false (default), filters out statements with non-gene bound conditions

**Returns** `stmts_out` – A list of filtered statements.

**Return type** `list[indra.statements.Statement]`

`indra.tools.assemble_corpus.filter_grounded_only(stmts_in, **kwargs)`

Filter to statements that have grounded agents.

#### Parameters

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **score\_threshold** (*Optional[float]*) – If scored groundings are available in a list and the highest score is below this threshold, the Statement is filtered out.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.
- **remove\_bound** (*Optional[bool]*) – If true, removes ungrounded bound conditions from a statement. If false (default), filters out statements with ungrounded bound conditions.

**Returns** `stmts_out` – A list of filtered statements.

**Return type** `list[indra.statements.Statement]`

`indra.tools.assemble_corpus.filter_human_only(stmts_in, **kwargs)`

Filter out statements that are grounded, but not to a human gene.

#### Parameters

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.
- **remove\_bound** (*Optional[bool]*) – If true, removes all bound conditions that are grounded but not to human genes. If false (default), filters out statements with boundary conditions that are grounded to non-human genes.

**Returns** `stmts_out` – A list of filtered statements.

**Return type** `list[indra.statements.Statement]`

`indra.tools.assemble_corpus.filter_inconsequential_acts(stmts_in, whitelist=None, **kwargs)`

Filter out Activations that modify inconsequential activities

Inconsequential here means that the site is not mentioned / tested in any other statement. In some cases specific activity types should be preserved, for instance, to be used as readouts in a model. In this case, the given activities can be passed in a whitelist.

**Parameters**

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **whitelist** (*Optional[dict]*) – A whitelist containing agent activity types which should be preserved even if no other statement refers to them. The whitelist parameter is a dictionary in which the key is a gene name and the value is a list of activity types. Example: `whitelist = {'MAP2K1': ['kinase']}`
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.

**Returns** `stmts_out` – A list of filtered statements.

**Return type** `list[indra.statements.Statement]`

```
indra.tools.assemble_corpus.filter_inconsequential_mods(stmts_in, whitelist=None,
                                                         **kwargs)
```

Filter out Modifications that modify inconsequential sites

Inconsequential here means that the site is not mentioned / tested in any other statement. In some cases specific sites should be preserved, for instance, to be used as readouts in a model. In this case, the given sites can be passed in a whitelist.

**Parameters**

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **whitelist** (*Optional[dict]*) – A whitelist containing agent modification sites whose modifications should be preserved even if no other statement refers to them. The whitelist parameter is a dictionary in which the key is a gene name and the value is a list of tuples of (modification\_type, residue, position). Example: `whitelist = {'MAP2K1': [('phosphorylation', 'S', '222')}`
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.

**Returns** `stmts_out` – A list of filtered statements.

**Return type** `list[indra.statements.Statement]`

```
indra.tools.assemble_corpus.filter_mod_nokinase(stmts_in, **kwargs)
```

Filter non-phospho Modifications to ones with a non-kinase enzyme.

**Parameters**

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.

**Returns** `stmts_out` – A list of filtered statements.

**Return type** `list[indra.statements.Statement]`

```
indra.tools.assemble_corpus.filter_mutation_status(stmts_in, mutations, deletions,
                                                    **kwargs)
```

Filter statements based on existing mutations/deletions

This filter helps to contextualize a set of statements to a given cell type. Given a list of deleted genes, it removes statements that refer to these genes. It also takes a list of mutations and removes statements that refer to mutations not relevant for the given context.

**Parameters**

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **mutations** (*dict*) – A dictionary whose keys are gene names, and the values are lists of tuples of the form (residue\_from, position, residue\_to). Example: mutations = {'BRAF': [('V', '600', 'E')]}.
- **deletions** (*list*) – A list of gene names that are deleted.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.

**Returns** **stmts\_out** – A list of filtered statements.

**Return type** list[indra.statements.Statement]

`indra.tools.assemble_corpus.filter_no_hypothesis` (*stmts\_in, \*\*kwargs*)

Filter to statements that are not marked as hypothesis in epistemics.

#### Parameters

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.

**Returns** **stmts\_out** – A list of filtered statements.

**Return type** list[indra.statements.Statement]

`indra.tools.assemble_corpus.filter_no_negated` (*stmts\_in, \*\*kwargs*)

Filter to statements that are not marked as negated in epistemics.

#### Parameters

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.

**Returns** **stmts\_out** – A list of filtered statements.

**Return type** list[indra.statements.Statement]

`indra.tools.assemble_corpus.filter_top_level` (*stmts\_in, \*\*kwargs*)

Filter to statements that are at the top-level of the hierarchy.

Here top-level statements correspond to most specific ones.

#### Parameters

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.

**Returns** **stmts\_out** – A list of filtered statements.

**Return type** list[indra.statements.Statement]

`indra.tools.assemble_corpus.filter_transcription_factor` (*stmts\_in, \*\*kwargs*)

Filter out RegulateAmounts where subject is not a transcription factor.

#### Parameters

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (stmts\_out) into.

**Returns** `stmts_out` – A list of filtered statements.

**Return type** `list[indra.statements.Statement]`

`indra.tools.assemble_corpus.filter_uuid_list` (*stmts\_in*, *uuids*, *\*\*kwargs*)

Filter to Statements corresponding to given UUIDs

**Parameters**

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to filter.
- **uuids** (*list[str]*) – A list of UUIDs to filter for.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (`stmts_out`) into.
- **invert** (*Optional[bool]*) – Invert the filter to remove the Statements corresponding to the given UUIDs.

**Returns** `stmts_out` – A list of filtered statements.

**Return type** `list[indra.statements.Statement]`

`indra.tools.assemble_corpus.load_statements` (*fname*, *as\_dict=False*)

Load statements from a pickle file.

**Parameters**

- **fname** (*str*) – The name of the pickle file to load statements from.
- **as\_dict** (*Optional[bool]*) – If True and the pickle file contains a dictionary of statements, it is returned as a dictionary. If False, the statements are always returned in a list. Default: False

**Returns** `stmts` – A list or dict of statements that were loaded.

**Return type** `list`

`indra.tools.assemble_corpus.map_grounding` (*stmts\_in*, *\*\*kwargs*)

Map grounding using the GroundingMapper.

**Parameters**

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to map.
- **do\_rename** (*Optional[bool]*) – If True, Agents are renamed based on their mapped grounding.
- **grounding\_map** (*Optional[dict]*) – A user supplied grounding map which maps a string to a dictionary of database IDs (in the format used by Agents' `db_refs`).
- **misgrounding\_map** (*Optional[dict]*) – A user supplied misgrounding map which maps a string to a known misgrounding which can be eliminated by the grounding mapper.
- **ignores** (*Optional[list]*) – A user supplied list of ignorable strings which, if present as an Agent text in a Statement, the Statement is filtered out.
- **use\_adeft** (*Optional[bool]*) – If True, Adept will be attempted to be used for acronym disambiguation. Default: True
- **save** (*Optional[str]*) – The name of a pickle file to save the results (`stmts_out`) into.

**Returns** `stmts_out` – A list of mapped statements.

**Return type** `list[indra.statements.Statement]`

`indra.tools.assemble_corpus.map_sequence` (*stmts\_in*, *\*\*kwargs*)

Map sequences using the SiteMapper.

#### Parameters

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to map.
- **do\_methionine\_offset** (*boolean*) – Whether to check for off-by-one errors in site position (possibly) attributable to site numbering from mature proteins after cleavage of the initial methionine. If True, checks the reference sequence for a known modification at 1 site position greater than the given one; if there exists such a site, creates the mapping. Default is True.
- **do\_orthology\_mapping** (*boolean*) – Whether to check sequence positions for known modification sites in mouse or rat sequences (based on PhosphoSitePlus data). If a mouse/rat site is found that is linked to a site in the human reference sequence, a mapping is created. Default is True.
- **do\_isoform\_mapping** (*boolean*) – Whether to check sequence positions for known modifications in other human isoforms of the protein (based on PhosphoSitePlus data). If a site is found that is linked to a site in the human reference sequence, a mapping is created. Default is True.
- **use\_cache** (*boolean*) – If True, a cache will be created/used from the location specified by `SITEMAPPER_CACHE_PATH`, defined in your INDRA config or the environment. If False, no cache is used. For more details on the cache, see the SiteMapper class definition.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (`stmts_out`) into.

**Returns** `stmts_out` – A list of mapped statements.

**Return type** `list[indra.statements.Statement]`

`indra.tools.assemble_corpus.merge_deltas` (*stmts\_in*)

Gather and merge original Influence delta information from evidence.

This function is only applicable to Influence Statements that have subj and obj deltas. All other statement types are passed through unchanged. Polarities and adjectives for subjects and objects respectively are collected and merged by traversing all evidences of a Statement.

**Parameters** `stmts_in` (*list[indra.statements.Statement]*) – A list of INDRA Statements whose influence deltas should be merged. These Statements are meant to have been preassembled and potentially have multiple pieces of evidence.

**Returns** `stmts_out` – The list of Statements now with deltas merged at the Statement level.

**Return type** `list[indra.statements.Statement]`

`indra.tools.assemble_corpus.merge_groundings` (*stmts\_in*)

Gather and merge original grounding information from evidences.

Each Statement's evidences are traversed to find original grounding information. These groundings are then merged into an overall consensus grounding dict with as much detail as possible.

The current implementation is only applicable to Statements whose concept/agent roles are fixed. Complexes, Associations and Conversions cannot be handled correctly.

**Parameters** `stmts_in` (*list[indra.statements.Statement]*) – A list of INDRA Statements whose groundings should be merged. These Statements are meant to have been preassembled and potentially have multiple pieces of evidence.

**Returns** `stmts_out` – The list of Statements now with groundings merged at the Statement level.

**Return type** `list[indra.statements.Statement]`

`indra.tools.assemble_corpus.reduce_activities` (*stmts\_in*, *\*\*kwargs*)

Reduce the activity types in a list of statements

**Parameters**

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to reduce activity types in.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (*stmts\_out*) into.

**Returns** *stmts\_out* – A list of reduced activity statements.

**Return type** `list[indra.statements.Statement]`

`indra.tools.assemble_corpus.rename_db_ref` (*stmts\_in*, *ns\_from*, *ns\_to*, *\*\*kwargs*)

Rename an entry in the `db_refs` of each Agent.

This is particularly useful when old Statements in pickle files need to be updated after a namespace was changed such as ‘BE’ to ‘FPLX’.

**Parameters**

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements whose Agents’ `db_refs` need to be changed
- **ns\_from** (*str*) – The namespace identifier to replace
- **ns\_to** (*str*) – The namespace identifier to replace to
- **save** (*Optional[str]*) – The name of a pickle file to save the results (*stmts\_out*) into.

**Returns** *stmts\_out* – A list of Statements with Agents’ `db_refs` changed.

**Return type** `list[indra.statements.Statement]`

`indra.tools.assemble_corpus.run_preassembly` (*stmts\_in*, *\*\*kwargs*)

Run preassembly on a list of statements.

**Parameters**

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements to preassemble.
- **return\_toplevel** (*Optional[bool]*) – If True, only the top-level statements are returned. If False, all statements are returned irrespective of level of specificity. Default: True
- **poolsize** (*Optional[int]*) – The number of worker processes to use to parallelize the comparisons performed by the function. If None (default), no parallelization is performed. NOTE: Parallelization is only available on Python 3.4 and above.
- **size\_cutoff** (*Optional[int]*) – Groups with `size_cutoff` or more statements are sent to worker processes, while smaller groups are compared in the parent process. Default value is 100. Not relevant when parallelization is not used.
- **belief\_scorer** (*Optional[indra.belief.BeliefScorer]*) – Instance of `BeliefScorer` class to use in calculating Statement probabilities. If None is provided (default), then the default scorer is used.
- **hierarchies** (*Optional[dict]*) – Dict of hierarchy managers to use for preassembly
- **matches\_fun** (*function*) – A function to override the built-in `matches_key` function of statements.
- **refinement\_fun** (*function*) – A function to override the built-in `refinement_of` function of statements.



- **flatten\_evidence** (*Optional[bool]*) – If True, evidences are collected and flattened via supports/supported\_by links. Default: False
- **flatten\_evidence\_collect\_from** (*Optional[str]*) – String indicating whether to collect and flatten evidence from the *supports* attribute of each statement or the *supported\_by* attribute. If not set, defaults to ‘supported\_by’. Only relevant when *flatten\_evidence* is True.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (*stmts\_out*) into.
- **save\_unique** (*Optional[str]*) – The name of a pickle file to save the unique statements into.

**Returns** *stmts\_out* – A list of preassembled top-level statements.

**Return type** list[indra.statements.Statement]

`indra.tools.assemble_corpus.run_preassembly_duplicate` (*preassembler*, *beliefengine*, *\*\*kwargs*)

Run deduplication stage of preassembly on a list of statements.

#### Parameters

- **preassembler** (`indra.preassembler.Preassembler`) – A Preassembler instance
- **beliefengine** (`indra.belief.BeliefEngine`) – A BeliefEngine instance.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (*stmts\_out*) into.

**Returns** *stmts\_out* – A list of unique statements.

**Return type** list[indra.statements.Statement]

`indra.tools.assemble_corpus.run_preassembly_related` (*preassembler*, *beliefengine*, *\*\*kwargs*)

Run related stage of preassembly on a list of statements.

#### Parameters

- **preassembler** (`indra.preassembler.Preassembler`) – A Preassembler instance which already has a set of unique statements internally.
- **beliefengine** (`indra.belief.BeliefEngine`) – A BeliefEngine instance.
- **return\_toplevel** (*Optional[bool]*) – If True, only the top-level statements are returned. If False, all statements are returned irrespective of level of specificity. Default: True
- **poolsize** (*Optional[int]*) – The number of worker processes to use to parallelize the comparisons performed by the function. If None (default), no parallelization is performed. NOTE: Parallelization is only available on Python 3.4 and above.
- **size\_cutoff** (*Optional[int]*) – Groups with *size\_cutoff* or more statements are sent to worker processes, while smaller groups are compared in the parent process. Default value is 100. Not relevant when parallelization is not used.
- **flatten\_evidence** (*Optional[bool]*) – If True, evidences are collected and flattened via supports/supported\_by links. Default: False
- **flatten\_evidence\_collect\_from** (*Optional[str]*) – String indicating whether to collect and flatten evidence from the *supports* attribute of each statement or the *supported\_by* attribute. If not set, defaults to ‘supported\_by’. Only relevant when *flatten\_evidence* is True.

- **save** (*Optional[str]*) – The name of a pickle file to save the results (`stmts_out`) into.

**Returns** `stmts_out` – A list of preassembled top-level statements.

**Return type** `list[indra.statements.Statement]`

`indra.tools.assemble_corpus.standardize_names_groundings` (*stmts*)

Standardize the names of Concepts with respect to an ontology.

NOTE: this function is currently optimized for Influence Statements obtained from Eidos, Hume, Sofia and CWMS. It will possibly yield unexpected results for biology-specific Statements.

`indra.tools.assemble_corpus.strip_agent_context` (*stmts\_in, \*\*kwargs*)

Strip any context on agents within each statement.

**Parameters**

- **stmts\_in** (*list[indra.statements.Statement]*) – A list of statements whose agent context should be stripped.
- **save** (*Optional[str]*) – The name of a pickle file to save the results (`stmts_out`) into.

**Returns** `stmts_out` – A list of stripped statements.

**Return type** `list[indra.statements.Statement]`

## 4.10.2 Real-time feedback for assembly (`indra.tools.live_curation`)

This REST service allows real-time curation and belief updates for a corpus of INDRA Statements.

**class** `indra.tools.live_curation.Corpus` (*statements, raw\_statements=None, aws\_name='wm'*)

Represent a corpus of statements with curation.

**Parameters**

- **statements** (*list[indra.statement.Statement]*) – A list of INDRA Statements to embed in the corpus.
- **aws\_name** (*str*) – The name of the profile in the AWS credential file to use. ‘default’ is used by default.

**statements**

A dict of INDRA Statements keyed by UUID.

**Type** `dict`

**curations**

A dict keeping track of the curations submitted so far for Statement UUIDs in the corpus.

**Type** `dict`

**s3\_get** (*name, bucket='world-modelers', key\_base\_name='wm'*)

Fetch a corpus object from S3 in the form of three json files

The json files representing the object have S3 keys of the format `<key_base_name>/<name>/<file>.json`

**Parameters**

- **name** (*str*) – The name of the model to fetch. Is part of the S3 key.
- **bucket** (*str*) – The S3 bucket to fetch the Corpus from. Default: ‘world-modelers’.
- **key\_base\_name** (*str*) – The base object path to fetch the json files from. Is part of the S3 key. Default: ‘indra\_models’.

**s3\_put** (*name*, *bucket*='world-modelers', *key\_base\_name*='indra\_models')

Push a corpus object to S3 in the form of three json files

The json files representing the object have S3 keys of the format <key\_base\_name>/<name>/<file>.json

**Parameters**

- **name** (*str*) – The name of the model to upload. Is part of the S3 key.
- **bucket** (*str*) – The S3 bucket to upload the Corpus to. Default: 'world-modelers'.
- **key\_base\_name** (*str*) – The base object path to upload the json files to. Is part of the S3 key. Default: 'indra\_models'.

**Returns keys** – A tuple of three strings giving the S3 key to the pushed objects

**Return type** tuple(str)

**exception** `indra.tools.live_curation.InvalidCorpusError`

**class** `indra.tools.live_curation.LiveCurator` (*scorer=None*, *corpora=None*)

Class coordinating the real-time curation of a corpus of Statements.

**Parameters**

- **scorer** (`indra.belief.BeliefScorer`) – A scorer object to use for the curation
- **corpora** (`dict[str, Corpus]`) – A dictionary mapping corpus IDs to Corpus objects.

**get\_corpus** (*corpus\_id*)

Return a corpus given an ID.

If the corpus ID cannot be found, an `InvalidCorpusError` is raised.

**Parameters** **corpus\_id** (*str*) – The ID of the corpus to return.

**Returns** The corpus with the given ID.

**Return type** *Corpus*

**reset\_scorer** ()

Reset the scorer used for curation.

**submit\_curation** (*corpus\_id*, *curations*)

Submit correct/incorrect curations fo a given corpus.

**Parameters**

- **corpus\_id** (*str*) – The ID of the corpus to which the curations apply.
- **curations** (*dict*) – A dict of curations with keys corresponding to Statement UUIDs and values corresponding to correct/incorrect feedback.

**update\_beliefs** (*corpus\_id*)

Return updated belief scores for a given corpus.

**Parameters** **corpus\_id** (*str*) – The ID of the corpus for which beliefs are to be updated.

**Returns** A dictionary of belief scores with keys corresponding to Statement UUIDs and values to new belief scores.

**Return type** dict

### 4.10.3 Build a network from a gene list (`indra.tools.gene_network`)

**class** `indra.tools.gene_network.GeneNetwork` (*gene\_list*, *basename=None*,  
*bel\_corpus=None*)

Build a set of INDRA statements for a given gene list from databases.

#### Parameters

- **gene\_list** (*str*) – List of gene names.
- **basename** (*str or None (default)*) – Filename prefix to be used for caching of intermediates (Biopax OWL file, pickled statement lists, etc.). If `None`, no results are cached and no cached files are used.
- **bel\_corpus** (*str*) – Path to a BEL corpus to use. The default uses the BEL Large Corpus.

#### **gene\_list**

List of gene names

**Type** `str`

#### **basename**

Filename prefix for cached intermediates, or `None` if no cached used.

**Type** `str or None`

#### **results**

Dict containing results of preassembly (see return type for `run_preassembly()`).

**Type** `dict`

#### **get\_bel\_stmts** (*filter=False*)

Get relevant statements from the BEL large corpus.

Performs a series of neighborhood queries and then takes the union of all the statements. Because the query process can take a long time for large gene lists, the resulting list of statements are cached in a pickle file with the filename `<basename>_bel_stmts.pkl`. If the pickle file is present, it is used by default; if not present, the queries are performed and the results are cached.

**Parameters** **filter** (*bool*) – If `True`, includes only those statements that exclusively mention genes in `gene_list`. Default is `False`. Note that the full (unfiltered) set of statements are cached.

**Returns** List of INDRA statements extracted from the BEL large corpus.

**Return type** list of `indra.statements.Statement`

#### **get\_biopax\_stmts** (*filter=False, query='pathsbetween', database\_filter=None*)

Get relevant statements from Pathway Commons.

Performs a “paths between” query for the genes in `gene_list` and uses the results to build statements. This function caches two files: the list of statements built from the query, which is cached in `<basename>_biopax_stmts.pkl`, and the OWL file returned by the Pathway Commons Web API, which is cached in `<basename>_pc_pathsbetween.owl`. If these cached files are found, then the results are returned based on the cached file and Pathway Commons is not queried again.

#### Parameters

- **filter** (*Optional[bool]*) – If `True`, includes only those statements that exclusively mention genes in `gene_list`. Default is `False`.
- **query** (*Optional[str]*) – Defined what type of query is executed. The two options are ‘pathsbetween’ which finds paths between the given list of genes and only works if

more than 1 gene is given, and ‘neighborhood’ which searches the immediate neighborhood of each given gene. Note that for pathsbetween queries with more than 60 genes, the query will be executed in multiple blocks for scalability.

- **database\_filter** (*Optional[list[str]]*) – A list of PathwayCommons databases to include in the query.

**Returns** List of INDRA statements extracted from Pathway Commons.

**Return type** list of `indra.statements.Statement`

**get\_statements** (*filter=False*)

Return the combined list of statements from BEL and Pathway Commons.

Internally calls `get_biopax_stmts()` and `get_bel_stmts()`.

**Parameters** **filter** (*bool*) – If True, includes only those statements that exclusively mention genes in *gene\_list*. Default is False.

**Returns** List of INDRA statements extracted the BEL large corpus and Pathway Commons.

**Return type** list of `indra.statements.Statement`

**run\_preassembly** (*stmts, print\_summary=True*)

Run complete preassembly procedure on the given statements.

Results are returned as a dict and stored in the attribute *results*. They are also saved in the pickle file `<basename>_results.pkl`.

**Parameters**

- **stmts** (list of `indra.statements.Statement`) – Statements to preassemble.
- **print\_summary** (*bool*) – If True (default), prints a summary of the preassembly process to the console.

**Returns**

A dict containing the following entries:

- *raw*: the starting set of statements before preassembly.
- *duplicates1*: statements after initial de-duplication.
- *valid*: statements found to have valid modification sites.
- *mapped*: mapped statements (list of `indra.preassembler.sitemapper.MappedStatement`).
- *mapped\_stmts*: combined list of valid statements and statements after mapping.
- *duplicates2*: statements resulting from de-duplication of the statements in *mapped\_stmts*.
- *related2*: top-level statements after combining the statements in *duplicates2*.

**Return type** dict

#### 4.10.4 Build an executable model from a fragment of a large network (`indra.tools.executable_subnetwork`)

`indra.tools.executable_subnetwork.get_subnetwork` (*statements, nodes, relevance\_network=None, relevance\_node\_lim=10*)

Return a PySB model based on a subset of given INDRA Statements.

Statements are first filtered for nodes in the given list and other nodes are optionally added based on relevance in a given network. The filtered statements are then assembled into an executable model using INDRA's PySB Assembler.

**Parameters**

- **statements** (*list[indra.statements.Statement]*) – A list of INDRA Statements to extract a subnetwork from.
- **nodes** (*list[str]*) – The names of the nodes to extract the subnetwork for.
- **relevance\_network** (*Optional[str]*) – The UUID of the NDEX network in which nodes relevant to the given nodes are found.
- **relevance\_node\_lim** (*Optional[int]*) – The maximal number of additional nodes to add to the subnetwork based on relevance.

**Returns** **model** – A PySB model object assembled using INDRA's PySB Assembler from the INDRA Statements corresponding to the subnetwork.

**Return type** pysb.Model

#### 4.10.5 Build a model incrementally over time (`indra.tools.incremental_model`)

**class** `indra.tools.incremental_model.IncrementalModel` (*model\_fname=None*)

Assemble a model incrementally by iteratively adding new Statements.

**Parameters** **model\_fname** (*Optional[str]*) – The name of the pickle file in which a set of INDRA Statements are stored in a dict keyed by PubMed IDs. This is the state of an IncrementalModel that is loaded upon instantiation.

**stmts**

A dictionary of INDRA Statements keyed by PMIDs that stores the current state of the IncrementalModel.

**Type** dict[str, list[indra.statements.Statement]]

**assembled\_stmts**

A list of INDRA Statements after assembly.

**Type** list[indra.statements.Statement]

**add\_statements** (*pmid, stmts*)

Add INDRA Statements to the incremental model indexed by PMID.

**Parameters**

- **pmid** (*str*) – The PMID of the paper from which statements were extracted.
- **stmts** (*list[indra.statements.Statement]*) – A list of INDRA Statements to be added to the model.

**get\_model\_agents** ()

Return a list of all Agents from all Statements.

**Returns** **agents** – A list of Agents that are in the model.

**Return type** list[indra.statements.Agent]

**get\_statements** ()

Return a list of all Statements in a single list.

**Returns** **stmts** – A list of all the INDRA Statements in the model.

**Return type** list[indra.statements.Statement]

**get\_statements\_noprior** ()

Return a list of all non-prior Statements in a single list.

**Returns** **stmts** – A list of all the INDRA Statements in the model (excluding the prior).

**Return type** list[indra.statements.Statement]

**get\_statements\_prior** ()

Return a list of all prior Statements in a single list.

**Returns** **stmts** – A list of all the INDRA Statements in the prior.

**Return type** list[indra.statements.Statement]

**load\_prior** (*prior\_fname*)

Load a set of prior statements from a pickle file.

The prior statements have a special key in the stmts dictionary called “prior”.

**Parameters** **prior\_fname** (*str*) – The name of the pickle file containing the prior Statements.

**preassemble** (*filters=None, grounding\_map=None*)

Preassemble the Statements collected in the model.

Use INDRA’s GroundingMapper, Preassembler and BeliefEngine on the IncrementalModel and save the unique statements and the top level statements in class attributes.

Currently the following filter options are implemented: - **grounding**: require that all Agents in statements are grounded - **human\_only**: require that all proteins are human proteins - **prior\_one**: require that at least one Agent is in the prior model - **prior\_all**: require that all Agents are in the prior model

**Parameters**

- **filters** (*Optional[list[str]]*) – A list of filter options to apply when choosing the statements. See description above for more details. Default: None
- **grounding\_map** (*Optional[dict]*) – A user supplied grounding map which maps a string to a dictionary of database IDs (in the format used by Agents’ db\_refs).

**save** (*model\_fname='model.pkl'*)

Save the state of the IncrementalModel in a pickle file.

**Parameters** **model\_fname** (*Optional[str]*) – The name of the pickle file to save the state of the IncrementalModel in. Default: model.pkl

#### 4.10.6 The RAS Machine (`indra.tools.machine`)

##### Prerequisites

First, install the machine-specific dependencies:

```
pip install indra[machine]
```

##### Starting a New Model

To start a new model, run

```
python -m indra.tools.machine make model_name
```

Alternatively, the command line interface can be invoked with

```
indra-machine make model_name
```

where `model_name` corresponds to the name of the model to initialize.

This script generates the following folders and files

- `model_name`
- `model_name/log.txt`
- `model_name/config.yaml`
- `model_name/jsons/`

You should the edit `model_name/config.yaml` to set up the search terms and optionally the credentials to use Twitter, Gmail or NDEx bindings.

### Setting Up Search Terms

The `config.yml` file is a standard YAML configuration file. A template is available in `model_name/config.yaml` after having created the machine.

Two important fields in `config.yml` are `search_terms` and `search_genes` both of which are YAML lists. The entries of `search_terms` are used directly as queries in PubMed search (for more information on PubMed search strings, read [https://www.ncbi.nlm.nih.gov/books/NBK3827/#pubmedhelp.Searching\\_PubMed](https://www.ncbi.nlm.nih.gov/books/NBK3827/#pubmedhelp.Searching_PubMed)).

Example:

```
search_terms:  
- breast cancer  
- proteasome  
- apoptosis
```

The entries of `search_genes` is a special list in which `_only_` standard HGNC gene symbols are allowed. Entries in this list are also used to search PubMed but also serve as a list of *prior* genes that are known to be relevant for the model.

#Entries in this can be used to search #PubMed specifically for articles that are tagged with the gene's unique #identifier rather than its string name. This mode of searching for articles #on specific genes is much more reliable than searching for them using #string names.

Example:

```
search_genes:  
- AKT1  
- MAPK3  
- EGFR
```

### Extending a Model

To extend a model, run

```
python -m indra.tools.machine run_with_search model_name
```



Alternatively, the command line interface can be invoked with

```
indra-machine run_with_search model_name
```

Extending a model involves extracting PMIDs from emails (if Gmail credentials are given), and searching using INDRA's PubMed client with each entry of `search_terms` in `config.yaml` as a search term. INDRA's literature client is then used to find the full text corresponding to each PMID or its abstract when the full text is not available. The REACH parser is then used to read each new paper. INDRA uses the REACH output to construct Statements corresponding to mechanisms. It then adds them to an incremental model through a process of assembly involving duplication and overlap resolution and the application of filters.

```
indra.tools.machine.copy_default_config(destination)
```

Copies the default configuration to the given destination

**Parameters** `destination` (*str*) – The location to which a default RAS Machine config file is placed.

### 4.10.7 High-throughput Reading Tools (`indra.tools.reading`)

#### Remote Reading Pipelines

There are three reading pipelines that have been developed for reading on remote high-performance systems.

#### A pipeline that uses AWS Batch and caches on S3 (`indra.tools.reading.pmid_reading`)

This pipeline makes use of AWS Batch jobs to scale readings arbitrarily, and optimizes the reading by caching results on S3, thus preventing the user from needing to reread content unnecessarily. This pipeline may someday be retired in favor of the RDS reading pipeline (see below), however at this time this method is nominally maintained.

#### The machinery for reading a list of PMIDs with REACH or SPARSER (`indra.tools.reading.pmid_reading.read_pmids`)

```
indra.tools.reading.pmid_reading.read_pmids.get_stmts(pmid_unread, cleanup=True,
                                                       sparser_version=None)
```

Run sparser on the pmids in `pmids_unread`.

```
indra.tools.reading.pmid_reading.read_pmids.join_json_files(prefix)
```

Join different REACH output JSON files into a single JSON object.

The output of REACH is broken into three files that need to be joined before processing. Specifically, there will be three files of the form: `<prefix>.uaz.<subcategory>.json`.

**Parameters** `prefix` (*str*) – The absolute path up to the extensions that reach will add.

**Returns** `json_obj` – The result of joining the files, keyed by the three subcategories.

**Return type** dict

```
indra.tools.reading.pmid_reading.read_pmids.read_pmid(pmid, source, cont_path,
                                                       sparser_version, out-
                                                       buf=None, cleanup=True)
```

Run sparser on a single pmid.

```
indra.tools.reading.pmid_reading.read_pmid.run_reach (pmid_list,          base_dir,
                                                    num_cores,      start_index,
                                                    end_index,     force_read,
                                                    force_fulltext, cleanup=False,
                                                    verbose=True)
```

Run reach on a list of pmids.

```
indra.tools.reading.pmid_reading.read_pmid.run_sparsener (pmid_list,      tmp_dir,
                                                         num_cores,  start_index,
                                                         end_index,  force_read,
                                                         force_fulltext,
                                                         cleanup=True,      ver-
                                                        bose=True)
```

Run the sparsener reader on the pmids in pmid\_list.

### A pipeline that uses AWS Batch and RDS (`indra_db.reading`)

This pipeline no longer exists under the umbrella of INDRA, but rather lives in the INDRA DB repo:

[https://github.com/indralab/indra\\_db/tree/master/indra\\_db/reading](https://github.com/indralab/indra_db/tree/master/indra_db/reading)

More information can be found in the *indra\_db* documentation. Unlike the other pipelines, this system is aimed at continuous automatic reading of all content.

### A pipeline that uses StarCluster (`indra.tools.reading.starcluster_reading`)

The first pipeline developed for large scale reading. This method is no longer actively maintained.

### High-throughput reading CLI's

The reading tools include several Python CLI's to run reading or other related tasks.

#### Run reading on local files

Read a list of files located in your local directory.

```
usage: python -m indra.tools.reading.read_files [-h]
        [-r {reach,sparsener,trips} [{reach,
        ↪sparsener,trips} ...]]
        [-n N_PROC] [-s N_SAMP]
        [-I RANGE_STR] [-v] [-q] [-d]
        input_file output_name
```

#### Positional Arguments

- input\_file**            A file containing a list of files/file paths to be read. These should be nxml or txt files.
- output\_name**        Results will be pickled in files `<output_name>_stmts.pkl` and `<output_name>_readings.pkl`.

## Named Arguments

|                        |  |
|------------------------|--|
| <b>-r, --readers</b>   | Possible choices: reach, sparser, trips<br>List of readers to be used. |
| <b>-n, --num_procs</b> | Select the number of processes to use.<br>Default: 1                   |
| <b>-s, --sample</b>    | Read a random sample of size N_SAMP of the inputs.                     |
| <b>-I, --in_range</b>  | Only read input lines in the range given as <start>:<end>.             |
| <b>-v, --verbose</b>   | Include output from the readers.<br>Default: False                     |
| <b>-q, --quiet</b>     | Suppress most output. Overrides -v and -d options.<br>Default: False   |
| <b>-d, --debug</b>     | Set the logging to debug level.<br>Default: False                      |

## Run REACH and/or SPARSER locally on a list of PMIDs using S3 caching

Apply NLP readers to the content available for a list of pmids.

```
usage: python -m indra.tools.reading.pmid_reading.read_pmids
  [-h] [-r {reach,sparser,all}] [{reach,sparser,all} ...] [-u]
  [--force_fulltext] [--force_read] [-n NUM_CORES] [-v] [-m]
  [-s START_INDEX] [-e END_INDEX] [--shuffle] [-o OUT_DIR]
  basename pmid_list_file
```

## Positional Arguments

|                       |   |
|-----------------------|---|
| <b>basename</b>       | The name of this job.   |
| <b>pmid_list_file</b> | Path to a file containing a list of line separated pmids for the articles to be read. |

## Named Arguments

|                          |  |
|--------------------------|--|
| <b>-r, --reader</b>      | Possible choices: reach, sparser, all<br>Choose which reader(s) to use.<br>Default: ['all']                            |
| <b>-u, --upload_json</b> | Option to simply upload previously read json files. Overrides -r option, so no reading will be done.<br>Default: False |
| <b>--force_fulltext</b>  | Option to force reading of the full text.<br>Default: False  |

|                          |  |
|--------------------------|--|
| <b>--force_read</b>      | Option to force the reader to reread everything.<br>Default: False   |
| <b>-n, --num_cores</b>   | Select the number of cores you want to use.<br>Default: 1  |
| <b>-v, --verbose</b>     | Show more output to screen.<br>Default: False  |
| <b>-m, --messy</b>       | Choose to not clean up after run.<br>Default: True   |
| <b>-s, --start_index</b> | Select the first pmid in the list to start reading.<br>Default: 0  |
| <b>-e, --end_index</b>   | Select the last pmid in the list to read.  |
| <b>--shuffle</b>         | Select a random sample of the pmids provided. <code>-s/--start_index</code> will be ignored, and <code>-e/--end_index</code> will set the number of samples to take.<br>Default: False |
| <b>-o, --outdir</b>      | The output directory where stuff is written. This is only a temporary directory when reading. By default this will be the " <code>&lt;basename&gt;_out</code> ".                       |

### Submit AWS Batch reading jobs

Run reading by collecting content, and save as pickles. This option requires that ids are given as a list of pmids, one line per pmid.

```
usage: python -m indra.tools.reading.submit_reading_pipeline
       [-h] {read,combine,full} ...
```

### Job Type

|                 |  |
|-----------------|--|
| <b>job_type</b> | Possible choices: read, combine, full<br>Type of jobs to submit. |
|-----------------|--|

### Sub-commands:

#### read

Run reading on batch and cache INDRA Statements on S3.

```
python -m indra.tools.reading.submit_reading_pipeline read [-h]
                                                         [--start_ix START_IX]
                                                         [--end_ix END_IX]
                                                         [--force_read]
                                                         [--force_fulltext]
                                                         [--ids_per_job IDS_PER_JOB]
                                                         [-r {parser,reach,all} [
↳ {parser,reach,all} ...]]
```

(continues on next page)

(continued from previous page)

```
[--project PROJECT]
input_file basename
```

## Positional Arguments

|                   |   |
|-------------------|---|
| <b>input_file</b> | Path to file containing input ids of content to read. For the no-db options, this is simply a file with each line being a pmid. For the with-db options, this is a file where each line is of the form '<id type>:<id>', for example 'pmid:12345' |
| <b>basename</b>   | Defines job names and S3 keys   |

## Named Arguments

|                         |   |
|-------------------------|---|
| <b>--start_ix</b>       | Start index of ids to read.   |
| <b>--end_ix</b>         | End index of ids to read. If <i>None</i> , read content from all ids.                       |
| <b>--force_read</b>     | Read papers even if previously read by current REACH.<br>Default: False                     |
| <b>--force_fulltext</b> | Get full text content even if content already on S3.<br>Default: False                      |
| <b>--ids_per_job</b>    | Number of PMIDs to read for each AWS Batch job.<br>Default: 3000                            |
| <b>-r, --readers</b>    | Possible choices: sparser, reach, all<br>Choose which reader(s) to use.<br>Default: ['all'] |
| <b>--project</b>        | Set the project name. Default is DEFAULT_AWS_PROJECT in the config.                         |

## combine

Combine INDRA Statement subsets into a single file.

```
python -m indra.tools.reading.submit_reading_pipeline combine
[-h] [-r {sparser,reach,all} [{sparser,reach,all} ...]] [--project PROJECT]
basename
```

## Positional Arguments

|                 |                               |
|-----------------|-------------------------------|
| <b>basename</b> | Defines job names and S3 keys |
|-----------------|-------------------------------|

## Named Arguments

- r, --readers**      Possible choices: sparser, reach, all  
Choose which reader(s) to use.  
Default: ['all']
- project**          Set the project name. Default is DEFAULT\_AWS\_PROJECT in the config.

## full

Run reading and combine INDRA Statements when done.

```
python -m indra.tools.reading.submit_reading_pipeline full [-h]
                                                         [--start_ix START_IX]
                                                         [--end_ix END_IX]
                                                         [--force_read]
                                                         [--force_fulltext]
                                                         [--ids_per_job IDS_PER_JOB]
                                                         [-r {sparser,reach,all} [
↳{sparser,reach,all} ...]]
                                                         [--project PROJECT]
                                                         input_file basename
```

## Positional Arguments

- input\_file**          Path to file containing input ids of content to read. For the no-db options, this is simply a file with each line being a pmid. For the with-db options, this is a file where each line is of the form '<id type>:<id>', for example 'pmid:12345'
- basename**            Defines job names and S3 keys

## Named Arguments

- start\_ix**          Start index of ids to read.
- end\_ix**            End index of ids to read. If *None*, read content from all ids.
- force\_read**        Read papers even if previously read by current REACH.  
Default: False
- force\_fulltext**    Get full text content even if content already on S3.  
Default: False
- ids\_per\_job**        Number of PMIDs to read for each AWS Batch job.  
Default: 3000
- r, --readers**        Possible choices: sparser, reach, all  
Choose which reader(s) to use.  
Default: ['all']
- project**            Set the project name. Default is DEFAULT\_AWS\_PROJECT in the config.

Note that `python wait_for_complete.py ...` should be run as soon as this command completes successfully. For more details use `python wait_for_complete.py -h`.

## Monitor running AWS Batch jobs

Wait for a set of batch jobs to complete, and monitor them as they run.

```
usage: python -m indra.tools.reading.wait_for_complete [-h] queue_name [options]
```

### Positional Arguments

**queue\_name** The name of the queue to watch and wait for completion. If no jobs are specified, this will wait until all jobs in the queue are completed (either SUCCEEDED or FAILED).

### Named Arguments

**--watch, -w** Specify particular jobs using their job ids, as reported by the submit command. Many ids may be specified.

**--prefix, -p** Specify a prefix for the name of the jobs to watch and wait for.

**--interval, -i** The time interval to wait between job status checks, in seconds (default: 10 seconds).

Default: 10

**--timeout, -T** If the logs are not updated for TIMEOUT seconds, print a warning. If `--kill_on_log_timeout` flag is set, then the offending jobs will be automatically terminated.

**--kill\_on\_timeout, -K** If a log times out, terminate the offending job.

Default: False

**--stash\_log\_method, -l** Possible choices: s3, local

Select a method from: [['s3', 'local']] to store the job logs. If no method is specified, the logs will not be loaded off of AWS. If 's3' is specified, then `job_name_prefix` must also be given, as this will indicate where on s3 to store the logs.

Jobs can also be monitored, terminated, and otherwise managed on the AWS website. However this tool will also tag the instances, and should be run whenever a job is submitted to AWS.

## Run the DRUM reading system

Run DRUM reading on a file.

```
usage: python -m indra.tools.reading.run_drum_reading [-h] file_name host port
```

## Positional Arguments

|                  |  |
|------------------|--|
| <b>file_name</b> | The name of the file to be read.                 |
| <b>host</b>      | The host on which DRUM is running.               |
| <b>port</b>      | The port to which the DRUM process is listening. |

## Generate stats on AWS Batch reading results

Get statistics on a bunch of statements.

```
usage: python -m indra.tools.reading.util.reading_results_stats
       [-h] {from-pickle,from-db} ...
```

## Source

|               |  |
|---------------|--|
| <b>source</b> | Possible choices: from-pickle, from-db |
|---------------|--|

## Sub-commands:

### from-pickle

Get statistics of statements in a pickle file.

```
python -m indra.tools.reading.util.reading_results_stats from-pickle
[-h] file_path
```

## Positional Arguments

|                  |   |
|------------------|---|
| <b>file_path</b> | The path to the pickle file containing the statements to be analyzed. |
|------------------|---|

### from-db

Get statistics from statements on the database.

```
python -m indra.tools.reading.util.reading_results_stats from-db
[-h] [--indra_version INDRA_VERSION] [--date_range DATE_RANGE]
```

## Named Arguments

|                        |  |
|------------------------|--|
| <b>--indra_version</b> | Specify the indra version for the batch of statements  |
| <b>--date_range</b>    | Specify the range of datetimes for statements. Must be in the format: “YYYYM-MDDHHMMSS:YYMMDDHHMMSS”. If you do not want to impose the upper or lower bound, simply leave it blank, eg. “YYYYMMDDHHMMSS:” if you don’t care about the upper bound. |



## High-throughput reading tools (`indra.tools.reading`)

INDRA defines interfaces to many text reading tools, however many of those only handle reading at small scales. These tools are developed to harness reading at arbitrary scales.

### Tools used to run reading on a set of locally stored files (`indra.tools.reading.read_files`)

Read a list of files located in your local directory.

```
indra.tools.reading.read_files.make_parser()
```

Create the argument parser, derived from the general scripts parser.

```
indra.tools.reading.read_files.read_files(files, readers, **kwargs)
```

Read the files in *files* with the reader objects in *readers*.

#### Parameters

- **files** (*list [str]*) – A list of file paths to be read by the readers. Supported files are limited to text and nxml files.
- **readers** (*list [Reader instances]*) – A list of Reader objects to be used reading the files.
- **\*\*kwargs** – Other keyword arguments are passed to the *read* method of the readers.

**Returns** **output\_list** – A list of ReadingData objects with the contents of the readings.

**Return type** list [[ReadingData](#)]

### Classes defining and implementing interfaces to different readers (`indra.tools.reading.readers`)

Objects for interacting with bulk nlp reading tools.

```
class indra.tools.reading.readers.Content(id, format, compressed=False, encoded=False)
```

An object to regularize the content passed to the readers.

**To use this class, use one of the two constructor methods:**

- *from\_file* : use content from a file on the filesystem.
- *from\_string* : Pass a string (or bytes) directly as content.

This class also regularizes the handling of id's and formats, as well as allowing for decompression and decoding, in the manner standard in the INDRA project.

```
change_format(new_format)
```

Change the format label of this content.

Note that this does NOT actually alter the format of the content, only the label.

```
change_id(new_id)
```

Change the id of this content.

```
classmethod from_file(file_path, compressed=False, encoded=False)
```

Create a content object from a file path.

```
classmethod from_string(id, format, raw_content, compressed=False, encoded=False)
```

Create a Content object from string/bytes content.

**get\_filename** (*renew=False*)

Get the filename of this content.

If the file name doesn't already exist, we created it as {id}.{format}.

**get\_filepath** (*renew=False*)

Get the file path, joining the name and location for this file.

If no location is given, it is assumed to be "here", e.g. ".".

**get\_text** ()

Get the loaded, decompressed, and decoded text of this content.

**is\_format** (*\*formats*)

Check the format of this content.

**set\_location** (*new\_location*)

Set/change the location of this content.

Note that this does NOT change the actual location of the file. To do so, use the *copy\_to* method.

```
class indra.tools.reading.readers.EmptyReader (base_dir=None, n_proc=1,  
                                              check_content=True, in-  
                                              put_character_limit=500000.0,  
                                              max_space_ratio=0.5, Re-  
                                              sultClass=<class 'in-  
                                              dra.tools.reading.readers.ReadingData'>)
```

A class name to use for Readers that are not implemented yet.

**exception** indra.tools.reading.readers.**ReachError**

**class** indra.tools.reading.readers.**ReachReader** (*\*args, \*\*kwargs*)

This object encodes an interface to the reach reading script.

**clear\_input** ()

Remove all the input files (at the end of a reading).

**get\_output** ()

Get the output of a reading job as a list of filenames.

**prep\_input** (*read\_list*)

Apply the readers to the content.

**read** (*read\_list, verbose=False, log=False*)

Read the content, returning a list of ReadingData objects.

```
class indra.tools.reading.readers.Reader (base_dir=None, n_proc=1, check_content=True,  
                                          input_character_limit=500000.0,  
                                          max_space_ratio=0.5, ResultClass=<class  
                                          'indra.tools.reading.readers.ReadingData'>)
```

This abstract object defines and some general methods for readers.

**add\_result** (*content\_id, content, \*\*kwargs*)

"Add a result to the list of results.

**read** (*read\_list, verbose=False, log=False*)

Read a list of items and return a dict of output files.

```
class indra.tools.reading.readers.ReadingData (content_id, reader, reader_version, con-  
                                              tent_format, content)
```

Object to contain the data produced by a reading.

#### Parameters

- **content\_id** (*int or str*) – A unique identifier of the text content that produced the reading, which can be mapped back to that content.
- **reader** (*str*) – The name of the reader, consistent with its *name* attribute, for example: 'REACH'
- **reader\_version** (*str*) – A string identifying the version of the underlying nlp reader.
- **content\_format** (*str*) – The format of the content. Options are in `indra.db.formats`.
- **content** (*str or dict*) – The content of the reading result. A string in the format given by *content\_format*.

**get\_statements** (*reprocess=False*)  
General method to create statements.

**exception** `indra.tools.reading.readers.ReadingError`

**exception** `indra.tools.reading.readers.SparsersError`

**class** `indra.tools.reading.readers.SparsersReader` (*\*args, \*\*kwargs*)  
This object provides methods to interface with the commandline tool.

**get\_output** (*output\_files, clear=True*)  
Get the output files as an id indexed dict.

**prep\_input** (*read\_list*)  
Prepare the list of files or text content objects to be read.

**read** (*read\_list, verbose=False, log=False, n\_per\_proc=None*)  
Perform the actual reading.

**read\_some** (*fpath\_list, outbuf=None, verbose=False*)  
Perform a few readings.

**class** `indra.tools.reading.readers.TripsReader` (*\*args, \*\*kwargs*)  
A stand-in for TRIPS reading.

Currently, we do not run TRIPS (more specifically DRUM) regularly at large scales, however on occasion we have outputs from TRIPS that were generated a while ago.

**read** (*\*args, \*\*kwargs*)  
Read a list of items and return a dict of output files.

`indra.tools.reading.readers.get_reader` (*reader\_name, \*args, \*\*kwargs*)  
Get an instantiated reader by name.

`indra.tools.reading.readers.get_reader_class` (*reader\_name*)  
Get a particular reader class by name.

`indra.tools.reading.readers.get_reader_classes` (*parent=<class 'indra.tools.reading.readers.Reader'>*)  
Get all childless the descendants of a parent class, recursively.

## Tools to run the DRUM reading system (`indra.tools.reading.run_drum_reading`)

`indra.tools.reading.run_drum_reading.read_pmid_sentences` (*pmid\_sentences, \*\*drum\_args*)  
Read sentences from a PMID-keyed dictionary and return all Statements

### Parameters

- **pmid\_sentences** (*dict[str, list[str]]*) – A dictionary where each key is a PMID pointing to a list of sentences to be read.
- **\*\*drum\_args** – Keyword arguments passed directly to the DrumReader. Typical things to specify are *host* and *port*. If *run\_drum* is specified as True, this process will internally run the DRUM reading system as a subprocess. Otherwise, DRUM is expected to be running independently.

**Returns** **all\_statements** – A list of INDRA Statements resulting from the reading

**Return type** list[indra.statement.Statement]

`indra.tools.reading.run_drum_reading.read_text(text, **drum_args)`  
Read sentences from a PMID-keyed dictionary and return all Statements

**Parameters**

- **text** (*str*) – A block of text to run DRUM on
- **\*\*drum\_args** – Keyword arguments passed directly to the DrumReader. Typical things to specify are ‘host’ and ‘port’.

**Returns** **statements** – A list of INDRA Statements resulting from the reading

**Return type** list[indra.statement.Statement]

## Python tools for submitting reading pipelines (`indra.tools.reading.submit_reading_pipeline`)

**exception** `indra.tools.reading.submit_reading_pipeline.BatchReadingError`

`indra.tools.reading.submit_reading_pipeline.get_ecs_cluster_for_queue(queue_name, batch_client=None)`  
Get the name of the ecs cluster using the batch client.

`indra.tools.reading.submit_reading_pipeline.submit_combine(basename, readers, job_ids=None, project_name=None)`  
Submit a batch job to combine the outputs of a reading job.

This function is provided for backwards compatibility. You should use the `PmidSubmitter` and `submit_combine` methods.

`indra.tools.reading.submit_reading_pipeline.submit_reading(basename, pmid_list_filename, readers, start_ix=None, end_ix=None, pmids_per_job=3000, num_tries=2, force_read=False, force_fulltext=False, project_name=None)`  
Submit an old-style pmid-centered no-database s3 only reading job.

This function is provided for the sake of backward compatibility. It is preferred that you use the object-oriented `PmidSubmitter` and the `submit_reading` job going forward.

`indra.tools.reading.submit_reading_pipeline.tag_instances_on_cluster(cluster_name, project='cwc')`  
Adds project tag to untagged instances in a given cluster.

**Parameters**

- **cluster\_name** (*str*) – The name of the AWS ECS cluster in which running instances should be tagged.
- **project** (*str*) – The name of the project to tag instances with.

```
indra.tools.reading.submit_reading_pipeline.wait_for_complete(queue_name,
                                                            job_list=None,
                                                            job_name_prefix=None,
                                                            poll_interval=10,
                                                            idle_log_timeout=None,
                                                            kill_on_log_timeout=False,
                                                            stash_log_method=None,
                                                            tag_instances=False,
                                                            re-
                                                            sult_record=None,
                                                            wait_for_first_job=False)
```

Return when all jobs in the given list finished.

If not job list is given, return when all jobs in queue finished.

**Parameters**

- **queue\_name** (*str*) – The name of the queue to wait for completion.
- **job\_list** (*Optional[list(dict)]*) – A list of jobID-s in a dict, as returned by the submit function. Example: [{‘jobId’: ‘e6b00f24-a466-4a72-b735-d205e29117b4’}, ...] If not given, this function will return if all jobs completed.
- **job\_name\_prefix** (*Optional[str]*) – A prefix for the name of the jobs to wait for. This is useful if the explicit job list is not available but filtering is needed.
- **poll\_interval** (*Optional[int]*) – The time delay between API calls to check the job statuses.
- **idle\_log\_timeout** (*Optional[int] or None*) – If not None, then track the logs of the active jobs, and if new output is not produced after *idle\_log\_timeout* seconds, a warning is printed. If *kill\_on\_log\_timeout* is set to True, the job will also be terminated.
- **kill\_on\_log\_timeout** (*Optional[bool]*) – If True, and if *idle\_log\_timeout* is set, jobs will be terminated after timeout. This has no effect if *idle\_log\_timeout* is None. Default is False.
- **stash\_log\_method** (*Optional[str]*) – Select a method to store the job logs, either ‘s3’ or ‘local’. If no method is specified, the logs will not be loaded off of AWS. If ‘s3’ is specified, then *job\_name\_prefix* must also be given, as this will indicate where on s3 to store the logs.
- **tag\_instances** (*bool*) – Default is False. If True, apply tags to the instances. This is toady typically done by each job, so in most cases this should not be needed.
- **result\_record** (*dict*) – A dict which will be modified in place to record the results of the job
- **wait\_for\_first\_job** (*bool*) – Don’t exit until at least one job has been found. This is good if you are monitoring jobs that are submitted periodically, but can be a problem if there is a chance you might call this when no jobs will ever be run.

## Low-level utilities used for or by reading pipelines (`indra.tools.reading.util`)

General tools to help with bulk reading.

### Tools to analyze logs (`indra.tools.reading.util.log_analysis_tools`)

**exception** `indra.tools.reading.util.log_analysis_tools.GetReadingStatsError`

`indra.tools.reading.util.log_analysis_tools.analyze_db_reading` (*job\_prefix*, *reading\_queue='run\_db\_reading\_queue'*)

Run various analysis on a particular reading job.

`indra.tools.reading.util.log_analysis_tools.analyze_reach_log` (*log\_fname=None*, *log\_str=None*)

Return unfinished PMIDs given a log file name.

`indra.tools.reading.util.log_analysis_tools.get_logs_from_db_reading` (*job\_prefix*, *reading\_queue='run\_db\_reading\_queue'*)

Get the logs stashed on s3 for a particular reading.

`indra.tools.reading.util.log_analysis_tools.get_unyielding_tcids` (*log\_str*)

Extract the set of tcids for which no statements were created.

`indra.tools.reading.util.log_analysis_tools.separate_reach_logs` (*log\_str*)

Get the list of reach logs from the overall logs.

### Tools to export statements by reach rule (`indra.tools.reading.util.export_stmts_by_reach_rule`)

### Class for generating reports on a Batch reading job (`indra.tools.reading.util.reporter`)

**class** `indra.tools.reading.util.reporter.Reporter` (*name*)

A class to produce a simple pdf report using the reportlab package.

**Parameters** *name* (*str*) – A name that will be the name of this file, and the title by default.

**add\_image** (*image\_path*, *width=None*, *height=None*, *section=None*)

Add an image to the document.

Images are shown on the final document in the order they are added, either within the given section or as part of the un-sectioned list of content.

#### Parameters

- **image\_path** (*str*) – A path to the image on the local file system.
- **width** (*int or float*) – The width of the image in the document in inches.
- **height** (*int or float*) – The height of the image in the document in inches.
- **section** (*str*) – (This must be a keyword) Select a section in which to place this image. Default is None, in which case the image will be simply be added to a default list of text and images.

**add\_section** (*section\_name*)

Create a section of the report, to be headed by *section\_name*

Text and images can be added by using the *section* argument of the *add\_text* and *add\_image* methods. Sections can also be ordered by using the *set\_section\_order* method.

By default, text and images that have no section will be placed after all the sections, in the order they were added. This behavior may be altered using the `sections_first` attribute of the `make_report` method.

**add\_text** (*text*, \*args, \*\*kwargs)

Add text to the document.

Text is shown on the final document in the order it is added, either within the given section or as part of the un-sectioned list of content.

#### Parameters

- **text** (*str*) – The text to be added.
- **style** (*str*) – Choose the style of the text. Options include ‘Normal’, ‘Code’, ‘Title’, ‘h1’. For others, see `getSampleStyleSheet` from `reportlab.lib.styles`.
- **space** (*tuple (num spaces, font size)*) – The number and size of spaces to follow this section of text. Default is (1, 12).
- **fontsize** (*int*) – The integer font size of the text (e.g. 12 for 12 point font). Default is 12.
- **alignment** (*str*) – The alignment of the text. Options include ‘left’, ‘right’, and ‘center’. Default is ‘left’.
- **section** (*str*) – (This must be a keyword) Select a section in which to place this text. Default is None, in which case the text will be simply be added to a default list of text and images.

**make\_report** (*sections\_first=True, section\_header\_params=None*)

Create the pdf document with name `self.name + '.pdf'`.

#### Parameters

- **sections\_first** (*bool*) – If True (default), text and images with sections are presented first and un-sectioned content is appended afterward. If False, sectioned text and images will be placed before the sections.
- **section\_header\_params** (*dict or None*) – Optionally overwrite/extend the default formatting for the section headers. Default is None.

**set\_section\_order** (*section\_name\_list*)

Set the order of the sections, which are by default unordered.

Any unlisted sections that exist will be placed at the end of the document in no particular order.

**set\_title** (*title*)

Overwrite the default title with a string title of your choosing.

## Tools used in creating CLI scripts for reading (`indra.tools.reading.util.script_tools`)

Useful tools for reading scripts.

`indra.tools.reading.util.script_tools.get_parser` (*description, input\_desc*)

Get a parser that is generic to reading scripts.

#### Parameters

- **description** (*str*) – A description of the tool, usually about one line long.
- **input\_desc** (*str*) – A string describing the nature of the input file used by the reading tool.

**Returns parser** – An argument parser object, to which further arguments can be added.

**Return type** `argparse.ArgumentParser` instance

## 4.11 Util (`indra.util`)

### 4.11.1 Utilities for using AWS (`indra.util.aws`)

`indra.util.aws.dump_logs(job_queue='run_reach_queue', job_status='RUNNING')`

Write logs for all jobs with given the status to files.

`indra.util.aws.get_batch_command(command_list, project=None, purpose=None)`

Get the command appropriate for running something on batch.

`indra.util.aws.get_job_log(job_info, log_group_name='/aws/batch/job', write_file=True, verbose=False)`

Gets the Cloudwatch log associated with the given job.

#### Parameters

- **job\_info** (*dict*) – dict containing entries for ‘jobName’ and ‘jobId’, e.g., as returned by `get_jobs()`
- **log\_group\_name** (*string*) – Name of the log group; defaults to ‘/aws/batch/job’
- **write\_file** (*boolean*) – If True, writes the downloaded log to a text file with the filename ‘%s\_%s.log’ % (job\_name, job\_id)

**Returns** The event messages in the log, with the earliest events listed first.

**Return type** list of strings

`indra.util.aws.get_jobs(job_queue='run_reach_queue', job_status='RUNNING')`

Returns a list of dicts with jobName and jobId for each job with the given status.

`indra.util.aws.get_log_by_name(log_group_name, log_stream_name, out_file=None, verbose=True)`

Download a log given the log’s group and stream name.

#### Parameters

- **log\_group\_name** (*str*) – The name of the log group, e.g. /aws/batch/job.
- **log\_stream\_name** (*str*) – The name of the log stream, e.g. run\_reach\_jobdef/default/<UUID>

**Returns lines** – The lines of the log as a list.

**Return type** list[str]

`indra.util.aws.get_s3_file_tree(s3, bucket, prefix)`

Overcome s3 response limit and return NestedDict tree of paths.

The NestedDict object also allows the user to search by the ends of a path.

The tree mimics a file directory structure, with the leaf nodes being the full unbroken key. For example, ‘path/to/file.txt’ would be retrieved by

```
ret['path']['to']['file.txt']['key']
```

The NestedDict object returned also has the capability to get paths that lead to a certain value. So if you wanted all paths that lead to something called ‘file.txt’, you could use



```
ret.get_paths('file.txt')
```

For more details, see the NestedDict docs.

```
indra.util.aws.kill_all(job_queue, reason='None given', states=None, kill_list=None)
```

Terminates/cancels all jobs on the specified queue.

#### Parameters

- **job\_queue** (*str*) – The name of the Batch job queue on which you wish to terminate/cancel jobs.
- **reason** (*str*) – Provide a reason for the kill that will be recorded with the job’s record on AWS.
- **states** (*None or list[str]*) – A list of job states to remove. Possible states are ‘STARTING’, ‘RUNNABLE’, and ‘RUNNING’. If None, all jobs in all states will be ended (modulo the *kill\_list* below).
- **kill\_list** (*None or list[dict]*) – A list of job dictionaries (as returned by the submit function) that you specifically wish to kill. All other jobs on the queue will be ignored. If None, all jobs on the queue will be ended (modulo the above).

**Returns** **killed\_ids** – A list of the job ids for jobs that were killed.

**Return type** list[str]

```
indra.util.aws.tag_instance(instance_id, **tags)
```

Tag a single ec2 instance.

```
indra.util.aws.tag_myself(project='cwc', **other_tags)
```

Function run when indra is used in an EC2 instance to apply tags.

### 4.11.2 A utility to get the INDRA version (`indra.util.get_version`)

This tool provides a uniform method for creating a robust indra version string, both from within python and from commandline. If possible, the version will include the git commit hash. Otherwise, the version will be marked with ‘UNHASHED’.

```
indra.util.get_version.get_git_info()
```

Get a dict with useful git info.

```
indra.util.get_version.get_version(with_git_hash=True, refresh_hash=False)
```

Get an indra version string, including a git hash.

### 4.11.3 Define NestedDict (`indra.util.nested_dict`)

```
class indra.util.nested_dict.NestedDict
```

A dict-like object that recursively populates elements of a dict.

More specifically, this acts like a recursive defaultdict, allowing, for example:

```
>> nd = NestedDict() >> nd['a']['b']['c'] = 'foo'
```

In addition, useful methods have been defined that allow the user to search the data structure. Note that the are not particularly optimized methods at this time. However, for convenience, you can for example simply call `get_path` to get the path to a particular key:

```
>> nd.get_path('c') (('a', 'b', 'c'), 'foo')
```

and the value at that key. Similarly:

```
>> nd.get_path('b') (('a', 'b'), NestedDict(
    'c': 'foo'
))
```

*get*, *gets*, and *get\_paths* operate on similar principles, and are documented below.

**export\_dict** ()

Convert this into an ordinary dict (of dicts).

**get** (*key*)

Find the first value within the tree which has the key.

**get\_leaves** ()

Get the deepest entries as a flat set.

**get\_path** (*key*)

Like *get*, but also return the path taken to the value.

**get\_paths** (*key*)

Like *gets*, but include the paths, like *get\_path* for all matches.

**gets** (*key*)

Like *get*, but return all matches, not just the first.

#### 4.11.4 Some shorthands for plot formatting (`indra.util.plot_formatting`)

```
indra.util.plot_formatting.format_axis(ax, label_padding=2, tick_padding=0, yticks_position='left')
```

Set standardized axis formatting for figure.

```
indra.util.plot_formatting.set_fig_params()
```

Set standardized font properties for figure.

## 5.1 Using natural language to build models

In this tutorial we build a simple model using natural language, and export it into different formats.

### 5.1.1 Read INDRA Statements from a natural language string

First we import INDRA's API to the TRIPS reading system. We then define a block of text which serves as the description of the mechanism to be modeled in the *model\_text* variable. Finally, *indra.sources.trips.process\_text* is called which sends a request to the TRIPS web service, gets a response and processes the extraction knowledge base to obtain a list of INDRA Statements

```
In [1]: from indra.sources import trips
In [2]: model_text = 'MAP2K1 phosphorylates MAPK1 and DUSP6 dephosphorylates MAPK1.'
In [3]: tp = trips.process_text(model_text)
```

At this point *tp.statements* should contain 2 INDRA Statements: a Phosphorylation Statement and a Dephosphorylation Statement. Note that the evidence sentence for each Statement is propagated:

```
In [4]: for st in tp.statements:
...:     print('%s with evidence "%s"' % (st, st.evidence[0].text))
...:
Phosphorylation(MAP2K1(), MAPK1()) with evidence "MAP2K1 phosphorylates MAPK1 and_
↪DUSP6 dephosphorylates MAPK1."
Dephosphorylation(DUSP6(), MAPK1()) with evidence "MAP2K1 phosphorylates MAPK1 and_
↪DUSP6 dephosphorylates MAPK1."
```

## 5.1.2 Assemble the INDRA Statements into a rule-based executable model

We next use INDRA's PySB Assembler to automatically assemble a rule-based model representing the biochemical mechanisms described in *model\_text*. First a PysbAssembler object is instantiated, then the list of INDRA Statements is added to the assembler. Finally, the assembler's *make\_model* method is called which assembles the model and returns it, while also storing it in *pa.model*. Notice that we are using *policies='two\_step'* as an argument of *make\_model*. This directs the assemble to use rules in which enzymatic catalysis is modeled as a two-step process in which enzyme and substrate first reversibly bind and the enzyme-substrate complex produces and releases a product irreversibly.

```
In [5]: from indra.assemblers.pysb import PysbAssembler

In [6]: pa = PysbAssembler()

In [7]: pa.add_statements(tp.statements)

In [8]: pa.make_model(policies='two_step')
Out[8]: <Model 'indra_model' (monomers: 3, rules: 6, parameters: 9, expressions: 0,
↳ compartments: 0) at 0x7fc12744d048>
```

At this point *pa.model* contains a PySB model object with 3 monomers,

```
In [9]: for monomer in pa.model.monomers:
...:     print(monomer)
...:
Monomer('MAP2K1', ['mapk'])
Monomer('MAPK1', ['phospho', 'map2k', 'dusp'], {'phospho': ['u', 'p']})
Monomer('DUSP6', ['mapk'])
```

6 rules,

```
In [10]: for rule in pa.model.rules:
...:     print(rule)
...:
Rule('MAP2K1_phosphorylation_bind_MAPK1_phospho', MAP2K1(mapk=None) + MAPK1(phospho='u
↳', map2k=None) >> MAP2K1(mapk=1) % MAPK1(phospho='u', map2k=1), kf_mm_bind_1)
Rule('MAP2K1_phosphorylation_MAPK1_phospho', MAP2K1(mapk=1) % MAPK1(phospho='u',
↳ map2k=1) >> MAP2K1(mapk=None) + MAPK1(phospho='p', map2k=None), kc_mm_
↳ phosphorylation_1)
Rule('MAP2K1_dissoc_MAPK1', MAP2K1(mapk=1) % MAPK1(map2k=1) >> MAP2K1(mapk=None) +
↳ MAPK1(map2k=None), kr_mm_bind_1)
Rule('DUSP6_dephosphorylation_bind_MAPK1_phospho', DUSP6(mapk=None) + MAPK1(phospho='p
↳', dusp=None) >> DUSP6(mapk=1) % MAPK1(phospho='p', dusp=1), kf_dm_bind_1)
Rule('DUSP6_dephosphorylation_MAPK1_phospho', DUSP6(mapk=1) % MAPK1(phospho='p',
↳ dusp=1) >> DUSP6(mapk=None) + MAPK1(phospho='u', dusp=None), kc_dm_phosphorylation_
↳ 1)
Rule('DUSP6_dissoc_MAPK1', DUSP6(mapk=1) % MAPK1(dusp=1) >> DUSP6(mapk=None) +
↳ MAPK1(dusp=None), kr_dm_bind_1)
```

and 9 parameters (6 kinetic rate constants and 3 total protein amounts) that are set to nominal but plausible values,

```
In [11]: for parameter in pa.model.parameters:
...:     print(parameter)
...:
Parameter('kf_mm_bind_1', 1e-06)
Parameter('kr_mm_bind_1', 0.1)
Parameter('kc_mm_phosphorylation_1', 100.0)
Parameter('kf_dm_bind_1', 1e-06)
```

(continues on next page)

(continued from previous page)

```
Parameter('kr_dm_bind_1', 0.1)
Parameter('kc_dm_phosphorylation_1', 100.0)
Parameter('MAP2K1_0', 10000.0)
Parameter('MAPK1_0', 10000.0)
Parameter('DUSP6_0', 10000.0)
```

The model also contains extensive annotations that tie the monomers to database identifiers and also annotate the semantics of each component of each rule.

```
In [12]: for annotation in pa.model.annotations:
        ....:     print(annotation)
        ....:
Annotation(MAP2K1, 'http://identifiers.org/hgnc/HGNC:6840', 'is')
Annotation(MAP2K1, 'http://identifiers.org/uniprot/Q02750', 'is')
Annotation(MAP2K1, 'http://identifiers.org/ncit/C17808', 'is')
Annotation(MAPK1, 'http://identifiers.org/hgnc/HGNC:6871', 'is')
Annotation(MAPK1, 'http://identifiers.org/uniprot/P28482', 'is')
Annotation(MAPK1, 'http://identifiers.org/ncit/C17589', 'is')
Annotation(DUSP6, 'http://identifiers.org/hgnc/HGNC:3072', 'is')
Annotation(DUSP6, 'http://identifiers.org/uniprot/Q16828', 'is')
Annotation(DUSP6, 'http://identifiers.org/ncit/C106024', 'is')
Annotation(MAP2K1_phosphorylation_bind_MAPK1_phospho, 'f0c0f86d-49ee-4124-b4c8-
↪383a8321536a', 'from_indra_statement')
Annotation(MAP2K1_phosphorylation_MAPK1_phospho, 'MAP2K1', 'rule_has_subject')
Annotation(MAP2K1_phosphorylation_MAPK1_phospho, 'MAPK1', 'rule_has_object')
Annotation(MAP2K1_phosphorylation_MAPK1_phospho, 'f0c0f86d-49ee-4124-b4c8-383a8321536a
↪', 'from_indra_statement')
Annotation(MAP2K1_dissoc_MAPK1, 'f0c0f86d-49ee-4124-b4c8-383a8321536a', 'from_indra_
↪statement')
Annotation(DUSP6_dephosphorylation_bind_MAPK1_phospho, 'e80510cb-ab4d-488d-8f2d-
↪6e8ba8937eff', 'from_indra_statement')
Annotation(DUSP6_dephosphorylation_MAPK1_phospho, 'DUSP6', 'rule_has_subject')
Annotation(DUSP6_dephosphorylation_MAPK1_phospho, 'MAPK1', 'rule_has_object')
Annotation(DUSP6_dephosphorylation_MAPK1_phospho, 'e80510cb-ab4d-488d-8f2d-
↪6e8ba8937eff', 'from_indra_statement')
Annotation(DUSP6_dissoc_MAPK1, 'e80510cb-ab4d-488d-8f2d-6e8ba8937eff', 'from_indra_
↪statement')
```

### 5.1.3 Exporting the model into other common formats

From the assembled PySB format it is possible to export the model into other common formats such as SBML, BNGL and Kappa. One can also generate a Matlab or Mathematica script with ODEs corresponding to the model.

```
pa.export_model('sbml')
pa.export_model('bngl')
```

One can also pass a file name argument to the *export\_model* function to save the exported model directly into a file:

```
pa.export_model('sbml', 'example_model.sbml')
```

## 5.2 The Statement curation interface

You will usually access this interface from an INDRA application that exposes statements to you. However if you just want to try out the interface or don't want to take the detour through any of the applications, you can follow the format below to access the interface directly in your browser from the INDRA-DB REST API:

```
http://api.host/statements/from_agents?subject=SUBJ&object=OBJ&api_key=12345&
format=html
```

where *api.host* should be replaced with the address to the REST API service (see the [documentation](#)). Entering the whole address in your browser will query for statements where *SUBJ* is the subject and *OBJ* is the object of the statements.

For more details about what options are available when doing curation, please refer to the [curation section](#) of the documentation.

### 5.2.1 Curating a Statement

Let's assume you want to check any statements where ROS1 is an agent for errors. Let's also limit the number of statements to 100 and the number of evidences per statements to 5. This will speed up the query and page loading. The appropriate address to enter in your browser would then be:

```
http://api.host/statements/from_agents?agent=ROS1&format=html&ev_limit=5&max_stmts=100
```

To start curating a statement, **click the pen icon (circled)** on the far left side of the statement. This will produce a row below the statement with a dropdown menu, a text box and a submit button:

The screenshot shows the 'INDRA DB Query Results' page. At the top, there is a 'Curator Information' section with fields for 'API key' (containing '123456') and 'Curator ID' (with a prompt to enter name or email). Below this, the statement 'ROS1 activates apoptosis' is displayed with a '47/4881' indicator. A 'Back to index' link is present. The main section is titled 'Evidence' and contains a table with columns for 'Source', 'Evidence', and 'Source Refs'. The first row of evidence is circled in blue, and a pen icon is also circled in blue. Below the evidence table, there is a 'Select error type...' dropdown menu, an 'Optional description (240 chars)' text box, and a 'Submit' button.

The **dropdown menu** contains common errors and also the possibility to mark the statement as 'correct'. If none of the types fit, select the *other...* option, and describe the error with one or a few words in the provided textbox. Note that if you pick *other...*, describing the error is mandatory. In our example, we see that *reactive oxygen species* is incorrectly grounded to *ROS*, so we pick *grounding* from the dropdown menu:

This screenshot shows the same interface as the previous one, but with the 'Select error type...' dropdown menu open. The 'grounding' option is selected and highlighted in orange. The 'Optional description' text box is empty. The 'Submit' button is visible. The evidence table below shows the same row as before, but with the 'grounding' error type applied to the evidence text.

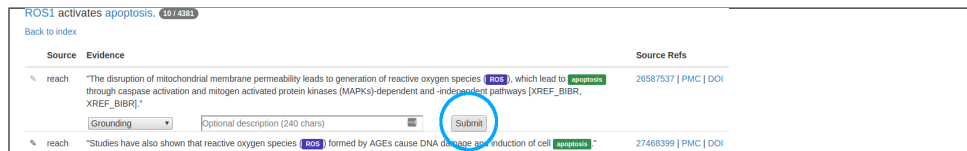
In the textbox, you can add a short optional description to clarify why you marked this piece of evidence with the error type you chose. When you are done, you are ready to submit your curation.

## 5.2.2 Submitting a Curation

To **submit a curation**, you will need to at least make a **selection in the dropdown menu** (by the curated statement). You will also need to be logged in before the curation is submitted. If you do not already have an account, all we ask for is your email.

If you selected *other...* in the dropdown menu, you must *also* describe the error in the textbox.

When you have entered the necessary information, click the **Submit button** by the statement that you curated (if you aren't logged in, you will be prompted to do so at this point):



A status message will appear once the server has processed the submission, indicating if the submission was successful or which problem arose if not. The pen icon will also change color based in the returned status. **Green** indicates a successful submission:

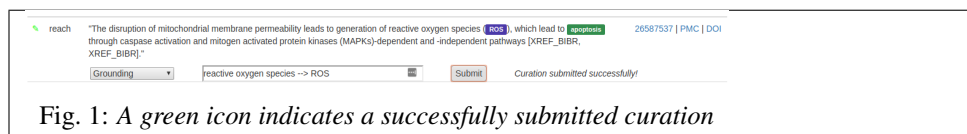


Fig. 1: A green icon indicates a successfully submitted curation

while a **red** indicates something went wrong with the submission:



Fig. 2: A red icon indicates that something went wrong during the submission

## 5.2.3 Curation Guidelines

### Basic principles

The main question to ask when deciding whether a given Statement is correct with respect to a given piece of evidence is:

Is there support in the evidence sentence for the Statement?

If the answer is **Yes**, then the given sentence is a valid piece of evidence for the Statement. In fact, you can assert this correctness by choosing the “Correct” option from the curation drop-down list. Curations that assert correctness are just as valuable as curations of incorrectness so the use of this option is encouraged.

Assuming the answer to the above question is **No**, one needs to determine what the error can be attributed to. The following section describes the specific error types that can be flagged.

### Types of errors to curate

There are currently the following options to choose from when curating incorrect Statement-sentence relationships:

- **Entity Boundaries:** this is applicable if the boundaries of one of the named entities was incorrectly recognized. Example: “gap” is highlighted as an entity, when in fact, the entity mentioned in the sentence was “gap junction”. These errors in entity boundaries almost always result in incorrect grounding, since the wrong string is attempted to be grounded. Therefore this error “subsumes” grounding errors. Note: to help correct entity boundaries, add the following to the Optional description text box: [gap junction], i.e. the desired entity name inside square brackets.

- **Grounding:** this is applicable if a named entity is assigned an incorrect database identifier. Example:

```
Assume that in a sentence, "ER" is mentioned referring to endoplasmic
reticulum, but in a Statement extracted from the sentence, it is
grounded to the ESR1 (estrogen receptor alpha) gene.
```

Note: to help correct grounding, add the following to the Optional description text box:

```
[ER] -> MESH:D004721
```

where [ER] is the entity string, MESH is the namespace of a database/ontology, and D004721 is the unique ID corresponding to endoplasmic reticulum in MESH. A list of commonly used namespaces in INDRA are given in: <https://indra.readthedocs.io/en/latest/modules/statements.html>. Note that you can also add multiple groundings separated by “|”, e.g. HGNC:11998|UP:P04637.

- **Polarity:** this is applicable if an essentially correct Statement was extracted but the Statement has the wrong polarity, e.g. Activation instead of Inhibition, of Phosphorylation instead of Dephosphorylation. Example:

```
Sentence: "NDRG2 overexpression specifically inhibits SOCS1 phosphorylation"
Statement: Phosphorylation(NDRG2(), SOCS1())
```

has incorrect polarity. It should be Dephosphorylation instead of Phosphorylation.

- **No Relation:** this is applicable if the sentence does not imply a relationship between the agents appearing in the Statement. Example:

```
Sentence: "Furthermore, triptolide mediated inhibition of NF-kappaB
activation, Stat3 phosphorylation and increase of SOCS1 expression in
DC may be involved in the inhibitory effect of triptolide."
Statement: Phosphorylation(STAT3(), SOCS1())
```

can be flagged as No Relation.

- **Wrong Relation Type:** this is applicable if the sentence implies a relationship between agents appearing in the Statement but the type of Statement is inconsistent with the sentence. Example:

```
Sentence: "We report the interaction between tacrolimus and chloramphenicol
in a renal transplant recipient."
Statement: Complex(tacrolimus(), chloramphenicol())
```

can be flagged as Wrong Relation Type since the sentence implies a drug interaction that does not involve complex formation.

- **Activity vs. Amount:** this is applicable when the sentence implies a regulation of amount but the corresponding Statement implies regulation of activity or vice versa. Example:

```
Sentence: "NFAT upregulates IL2"
Statement: Activation(NFAT(), IL2())
```

Here the sentence implies upregulation of the amount of IL2 but the corresponding Statement is of type Activation rather than IncreaseAmount.



- **Negative Result:** this is applicable if the sentence implies the lack of or opposite of a relationship. Example:

```
Sentence: "These results indicate that CRAF, but not BRAF phosphorylates
MEK in NRAS mutated cells."
Statement: Phosphorylation(BRAF(), MEK())
```

Here the sentence does not support the Statement due to a negation and should therefore be flagged as a Negative Result.

- **Hypothesis:** this is applicable if the sentence describes a hypothesis or an experiment rather than a result or mechanism. Example:

```
Sentence: "We tested whether EGFR activates ERK."
Statement: Activation(EGFR(), ERK())
```

Here the sentence describes a hypothesis with respect to the Statement, and should therefore be flagged as a Hypothesis upon curation (unless of course the Statement already has a correct *hypothesis* flag).

- **Agent Conditions:** this is applicable if one of the Agents in the Statement is missing relevant conditions that are mentioned in the sentence, or has incorrect conditions attached to it. Example:

```
Sentence: "Mutant BRAF activates MEK"
Statement: Activation(BRAF(), MEK())
```

can be curated to be missing Agent conditions since the mutation on BRAF is not captured.

- **Modification Site:** this is applicable if an amino-acid site is missing or incorrect in a modification Statement. Example:

```
Sentence: "MAP2K1 phosphorylates MAPK1 at T185."
Statement: Phosphorylation(MAP2K1(), MAPK1())
```

Here the obvious modification site is missing from MAPK1.

- **Other:** this is an option you can choose whenever the problem isn't well captured by any of the more specific options. In this case you need to add a note to explain what the issue is.

### General notes on curation

- If you spot multiple levels of errors in a Statement-sentence pair, use the most relevant error type in the dropdown menu. E.g. if you see both a grounding error and a polarity error, you should pick the grounding error since a statement with a grounding error generally would not exist if the grounding was correct.
- If you still feel like multiple errors are appropriate for the curation, select a new error from the dropdown menu and make a new submission.
- Please be consistent in using your email address as your curator ID. Keeping track of who curated what helps us to faster track down issues with readers and the assembly processes that generate statements.

## 5.3 Large-Scale Machine Reading with Starcluster

The following doc describes the steps involved in reading a large numbers of papers in parallel on Amazon EC2 using REACH, caching the JSON output on Amazon S3, then processing the REACH output into INDRA Statements. Prerequisites for doing the following are:

- A cluster of Amazon EC2 nodes configured using Starcluster, with INDRA installed and in the PYTHONPATH

- An Amazon S3 bucket containing full text contents for papers, keyed by Pubmed ID (creation of this S3 repository will be described in another tutorial).

This tutorial goes through the individual steps involved before describing how all of them can be run through the use of a single submission script, `submit_reading_pipeline.py`.

Note also that the prerequisite installation steps can be streamlined by putting them in a setup script that can be re-run upon instantiating a new Amazon cluster or by using them to configure a custom Amazon EC2 AMI.

### 5.3.1 Install REACH

Install SBT. On an EC2 Linux machine, run the following lines (drawn from <http://www.scala-sbt.org/0.13/docs/Installing-sbt-on-Linux.html>):

```
echo "deb https://dl.bintray.com/sbt/debian /" | sudo tee -a /etc/apt/sources.list.d/
↪sbt.list
sudo apt-key adv --keyserver hkp://keyserver.ubuntu.com:80 --recv 642AC823
sudo apt-get update
sudo apt-get install sbt
```

Clone REACH from <https://github.com/clulab/reach>.

Add the following line to `reach/build.sbt`:

```
mainClass in assembly := Some("org.clulab.reach.ReachCLI")
```

This assigns `ReachCLI` as the main class.

Compile and assemble REACH. Note that the path to the `.ivy2` directory must be given. Use the assembly task to assemble a fat JAR containing all of the dependencies with the correct main class. Run the following from the directory containing the REACH `build.sbt` file (e.g., `/pmc/reach`):

```
sbt -Dsbt.ivy.home=/pmc/reach/.ivy2 compile
sbt -Dsbt.ivy.home=/pmc/reach/.ivy2 assembly
```

### 5.3.2 Install Amazon S3 support

Install boto3:

```
pip install boto3
```

---

**Note:** If using EC2, make sure to install boto3, jsonpickle, and Amazon credentials on all nodes, not just the master node.

---

Add Amazon credentials to access the S3 bucket. First create the `.aws` directory on the EC2 instance:

```
mkdir /home/sgeadmin/.aws
```

Then set up Amazon credentials, for example by copying from your local machine using `StarCluster`:

```
starcluster put mycluster ~/.aws/credentials /home/sgeadmin/.aws
```

### 5.3.3 Install other dependencies

```
pip install jsonpickle # Necessary to process JSON from S3
pip install --upgrade pyjnius # Necessary for REACH
export JAVA_HOME=/usr/lib/jvm/java-1.7.0-openjdk-amd64
```

### 5.3.4 Assemble a Corpus of PMIDs

The first step in large-scale reading is to put together a file containing relevant Pubmed IDs. The simplest way to do this is to use the Pubmed search API to find papers associated with particular gene names, biological processes, or other search terms.

For example, to assemble a list of papers for SOS2 curated in Entrez Gene that are available in the Pubmed Central Open Access subset:

```
In [1]: from indra.literature import *

# Pick an example gene
In [2]: gene = 'SOS2'

# Get a list of PMIDs for the gene
In [3]: pmids = pubmed_client.get_ids_for_gene(gene)

# Get the PMIDs that have XML in PMC
In [4]: pmids_oa_xml = pmc_client.filter_pmids(pmids, 'oa_xml')

# Write the results to a file
In [5]: with open('%s_pmids.txt' % gene, 'w') as f:
...:     for pmid in pmids_oa_xml:
...:         f.write('%s\n' % pmid)
...:
```

This creates a file, SOS2\_pmids.txt, containing the PMIDs that we will read with REACH.

### 5.3.5 Process the papers with REACH

The next step is to read the content of the papers with REACH in a parallelizable, high-throughput way. To do this, run the script `indra/tools/reading/run_reach_on_pmids.py`. If necessary update the lines at the top of the script with the REACH settings, e.g.:

```
cleanup = False
verbose = True
path_to_reach = '/pmc/reach/target/scala-2.11/reach-assembly-1.3.2-SNAPSHOT.jar'
reach_version = '1.3.2'
source_text = 'pmc_oa_xml'
force_read = False
```

The `reach_version` is important because it is used to determine whether the paper has already been read with this version of REACH (in which case it will be skipped), or if the REACH output needs to be updated. Alternatively, if you want to read all the papers regardless of whether they've been read before with the given version of REACH, set the `force_read` variable to `True`.

Next, create a top-level temporary directory to use during reading. This will be used to store the input files and the JSON output:

```
mkdir my_temp_dir
```

Run `run_reach_on_pmids.py`, passing arguments for the PMID list file, the temp directory, the number of cores to use on the machine, the PMID start index (in the PMID list file) and the end index. The start and end indices are used to subdivide the job into parallelizable chunks. If the end index is greater than the total number of PMIDs, it will process up to the last one in the list. For example:

```
python run_reach_on_pmids.py SOS2_pmids.txt my_temp_dir 8 0 10
```

This uses 8 cores to process the first ten papers listed in the file `SOS2_pmids.txt`. REACH will run, output the JSON files in the temporary directory, e.g. in `my_temp_dir/read_0_to_10_MSP6YI/output`, assemble the JSON files together, and upload the results to S3. If you attempt to process the files again with the same version of REACH, the script will detect that the JSON output from that version is already on S3 and skip those papers.

This can be submitted to run offline using the job scheduler on EC2 with, e.g.:

```
qsub -b y -cwd -V -pe orte 8 python run_reach_on_pmids.py SOS2_pmids.txt my_temp_dir_
↪8 0 10
```

---

**Note:** The number of cores requested in the `qsub` call (`'-pe orte 8'`) should match the number of cores passed to the `run_reach_on_pmids.py` script, which determines the number of threads that REACH will attempt to use (the third-to-last argument above). This should also match the total number of nodes on the Amazon EC2 node (e.g., 8 cores for `c3.2xlarge`). This way the job scheduler will schedule the job to run on all the cores of a single EC2 node, and REACH will use them all.

---

### 5.3.6 Extract INDRA Statements from the REACH output on S3

The script `indra/tools/reading/process_reach_from_s3.py` is used to extract INDRA Statements from the REACH output uploaded to S3 in the previous step. This process can also be parallelized by submitting chunks of papers to be processed by different cores. The INDRA statements for each chunk of papers are pickled and can be assembled into a single pickle file in a subsequent step.

Following the example above, run the following to process the REACH output for the SOS2 papers into INDRA statements. We'll do this in two chunks to show how the process can be parallelized and the statements assembled from multiple files:

```
python process_reach_from_s3.py SOS2_pmids.txt 0 5
python process_reach_from_s3.py SOS2_pmids.txt 5 10
```

The two runs create two different files for the results from the seven papers, `reach_stmts_0_5.pkl` (with statements from the first five papers) and `reach_stmts_5_7.pkl` (with statements from the last two). Note that the results are pickled as a dict (rather than a list), with PMIDs as keys and lists of Statements as values.

Of course, what we really want is a single file containing all of the statements for the entire corpus. To get this, run:

```
python assemble_reach_stmts.py reach_stmts_*.pkl
```

The results will be stored in `reach_stmts.pkl`.

### 5.3.7 Running the whole pipeline with one script

If you want to run the whole pipeline in one go, you can run the script `submit_reading_pipeline.py` (in `indra/tools/reading`) on a cluster of Amazon EC2 nodes. The script divides up the jobs evenly among the nodes and

cores. Usage:

```
python submit_reading_pipeline.py pmid_list tmp_dir num_nodes num_cores_per_node
```

For example if you have a cluster with 8 c3.8xlarge nodes with 32 VCPUs each, you would call it with:

```
python submit_reading_pipeline.py SOS2_pmids.txt my_tmp_dir 8 32
```

The script submits the jobs to the scheduler with appropriate dependencies such that the REACH reading step completes first, then the INDRA processing step, and then the final assembly into a single pickle file.

## 5.4 Large-Scale Machine Reading with Amazon Batch

The following doc describes the steps involved in reading a large numbers of papers in parallel on Amazon EC2 using REACH, caching the JSON output on Amazon S3, processing the REACH output into INDRA Statements, and then caching the statements also on S3. Prerequisites for doing the following are:

- An Amazon S3 bucket containing full text contents for papers, keyed by Pubmed ID (creation of this S3 repository will be described in another tutorial).
- Amazon AWS credentials for using AWS Batch.
- A corpus of PMIDs (see *Large-Scale Machine Reading with Starcluster* for information on how to assemble this)
- Optional: Elsevier text and data mining API key and institution key for subscriber access to Elsevier full text content.

### 5.4.1 How it Works

- The reading pipeline makes use of a Docker image that contains INDRA and all necessary dependencies, including REACH, Kappa, PySB, etc. The Docker file for this image is available at: [https://github.com/johnbachman/indra\\_docker](https://github.com/johnbachman/indra_docker).
- The INDRA Docker image is built by AWS Codebuild and pushed to Amazon’s EC2 Container Service (ECS), where it is available via the Repository URI:

```
292075781285.dkr.ecr.us-east-1.amazonaws.com/indra
```

- An AWS Batch *Compute Environment* named “run\_reach” is configured to use this Docker image for handling AWS jobs. This compute environment is configured to use only Spot instances with a maximum spot price of 40% of the on-demand price, and 16 vCPUs.
- An AWS *Job Queue*, “run\_reach\_queue”, is configured to use instances of the “run\_reach” Compute Environment.
- An AWS *Job Definition*, “run\_reach\_jobdef”, is configured to run in the “run\_reach\_queue”, and to use 16 vCPUs and 30GiB of RAM.
- Reading jobs are submitted by running the script:

```
python -m indra.tools.reading.submit_reading_pipeline_aws read [args]
```

which, given a list of PMIDs:

- Copies the PMID list to the key `reading_results/[job_name]/pmids` on Amazon S3

- Breaks the list up into chunks (e.g., of 3000 PMIDs) and submits an AWS Batch job for each (using the “run\_reach\_jobdef” definition as a template).
- The ECS instance created by the AWS Batch job runs the script `indra.tools.reading.run_reach_on_pmids_aws`, which:
  - Checks for cached content on Amazon S3
  - If the PMID has not been read by the current version of REACH, checks for content
  - If the content is not available, downloads the content using the INDRA literature client, and caches on S3
  - The content to be read is downloaded to the `/tmp` directory of the instance
  - REACH is run using the command-line interface (RunReachCLI), and configured to read the papers in the `/tmp` directory using all of the vCPUs on the instance
  - When done, the result REACH JSON in the output folder is uploaded to S3
  - The JSON for both the previously and newly read papers is processed in parallel to INDRA Statements
  - The resulting subset of statements for the given range of papers is cached on S3 at `reading_results/[job_name]/stmts/[start_ix]_[end_ix].pkl`. This set of statements takes the form of a pickled (protocol 3) Python dict with PMIDs as keys and lists of INDRA Statements as values.
  - In addition, information about the sources of content available for each PMID is cached for each PMID subset at `reading_results/[job_name]/content_types/[start_ix]_[end_ix].pkl`.
- When the reading jobs for each of the subsets of PMIDs have been completed and cached on S3, the final combined set of statements (and combined information on content sources) can be assembled using:

```
python -m indra.tools.reading.submit_reading_pipeline_aws combine [job_name]
```

- This script submits an AWS batch job for a machine with 1 vCPU but a large amount of memory (60GiB)
- The job runs the script `indra.tools.reading.assemble_reach_stmts_aws`, which unpickles the results from all of the PMID subsets, combines them, and stores them on S3
- The resulting files are obtainable from S3 at `reading_results/[job_name]/stmts.pkl` and `reading_results/[job_name]/content_types.pkl`.
- To run the entire pipeline, where the assembly of the combined set of statements is automatically performed after the reading step is completed, run:

```
python -m indra.tools.reading.submit_reading_pipeline_aws full [args]
```

## 5.5 Assembling everything known about a particular gene

Assume you are interested in collecting all mechanisms that a particular gene is involved in. Using INDRA, it is possible to collect everything curated about the gene in pathway databases and then read all the accessible literature discussing the gene of interest. This knowledge is aggregated as a set of INDRA Statements which can then be assembled into several different model and network formats and possibly shared online.

For the sake of example, assume that the gene of interest is TMEM173.

It is important to use the standard HGNC gene symbol of the gene throughout the example (this information is available on <http://www.genenames.org/> or <http://www.uniprot.org/>) - arbitrary synonyms will not work!

### 5.5.1 Collect mechanisms from PathwayCommons and the BEL Large Corpus

We first collect Statements from the PathwayCommons database via INDRA's BioPAX API and then collect Statements from the BEL Large Corpus via INDRA's BEL API.

```
from indra.tools.gene_network import GeneNetwork

gn = GeneNetwork(['TMEM173'])
biopax_stmts = gn.get_biopax_stmts()
bel_stmts = gn.get_bel_stmts()
```

at this point *biopax\_stmts* and *bel\_stmts* are two lists of INDRA Statements.

### 5.5.2 Collect a list of publications that discuss the gene of interest

We next use INDRA's literature client to find PubMed IDs (PMIDs) that discuss the gene of interest. To find articles that are annotated with the given gene, INDRA first looks up the Entrez ID corresponding to the gene name and then finds associated publications.

```
from indra import literature

pmids = literature.pubmed_client.get_ids_for_gene('TMEM173')
```

The variable *pmid* now contains a list of PMIDs associated with the gene.

### 5.5.3 Get the full text or abstract corresponding to the publications

Next we use INDRA's literature client to fetch the full text (if available) or the abstract corresponding to the PMIDs we have just collected.

```
from indra import literature

paper_contents = {}
for pmid in pmids:
    content, content_type = literature.get_full_text(pmid, 'pmid')
    paper_contents[pmid] = (content, content_type)
```

We now have a dictionary called *paper\_contents* which stores the content and the content type of each PMID we looked up.

### 5.5.4 Read the content of the publications

We next run the REACH reading system on the publications. Depending on the content type, different calls need to be made via INDRA's REACH API.

```
from indra import literature
from indra.sources import reach

read_offline = True

literature_stmts = []
for pmid, (content, content_type) in paper_contents.items():
    rp = None
```

(continues on next page)

(continued from previous page)

```

print('Reading %s' % PMID)
if content_type == 'abstract':
    rp = reach.process_text(content, citation=PMID, offline=read_offline)
elif content_type == 'pmc_oa_xml':
    rp = reach.process_nxml_str(content, offline=read_offline)
elif content_type == 'elsevier_xml':
    txt = literature.elsevier_client.extract_text(content)
    if txt:
        rp = reach.process_text(txt, citation=PMID, offline=read_offline)
if rp is not None:
    literature_stmts += rp.statements

```

The list *literature\_stmts* now contains the results of all the statements that were read.

### 5.5.5 Combine all statements and run pre-assembly

```

from indra.tools import assemble_corpus

stmts = biopax_stmts + bel_stmts + literature_stmts

stmts = assemble_corpus.map_grounding(stmts)
stmts = assemble_corpus.map_sequence(stmts)
stmts = assemble_corpus.run_preassembly(stmts)

```

At this point *stmts* contains a list of Statements collected with grounding, sequences having been mapped, duplicates combined and less specific variants of statements hidden. It is possible to run other filters on the results such as to keep only human genes, remove Statements with ungrounded genes, or to keep only certain types of interactions.

### 5.5.6 Assemble the statements into a network model

```

from indra.assemblers.cx import CxAssembler
from indra.databases import ndex_client

cxa = CxAssembler(stmts)
cx_str = cxa.make_model()

```

We can now upload this network to the Network Data Exchange (NDEX).

```

ndex_cred = {'user': 'myusername', 'password': 'xxx'}
network_id = ndex_client.create_network(cx_str, ndex_cred)
print(network_id)

```



Many functionalities of INDRA can be used via a REST API. This enables making use of INDRA's knowledge sources and assembly capabilities in a RESTful, platform independent fashion. The REST service is meant to be used locally (on a single machine or local network) and is currently not offered as a public web service by the creators of INDRA.

### 6.1 Installation

The REST service requires the *bottle* package to be installed in addition to all the other requirements of INDRA.

### 6.2 Launching the REST service

The REST service can be launched by running *api.py* in the *rest\_api* folder within *indra*.

### 6.3 Documentation

The specific end-points and input/output parameters offered by the REST API are documented in *rest\_api/docs/index.html*, which is accessible locally within the *indra* folder.



## CHAPTER 7

---

### Indices and tables

---

- genindex
- modindex
- search



### i

indra.assemblers.cag.assembler, 149  
indra.assemblers.cx.assembler, 141  
indra.assemblers.cyjs.assembler, 148  
indra.assemblers.english.assembler, 142  
indra.assemblers.graph.assembler, 144  
indra.assemblers.html.assembler, 151  
indra.assemblers.index\_card.assembler, 146  
indra.assemblers.indranet, 158  
indra.assemblers.indranet.assembler, 160  
indra.assemblers.indranet.net, 158  
indra.assemblers.kami.assembler, 157  
indra.assemblers.pybel.assembler, 156  
indra.assemblers.pysb.assembler, 135  
indra.assemblers.pysb.base\_agents, 140  
indra.assemblers.pysb.bmi\_wrapper, 153  
indra.assemblers.pysb.kappa\_util, 141  
indra.assemblers.pysb.preassembler, 140  
indra.assemblers.sbgm.assembler, 147  
indra.assemblers.sif.assembler, 145  
indra.assemblers.tsv.assembler, 150  
indra.belief, 129  
indra.databases.biogrid\_client, 93  
indra.databases.cbio\_client, 96  
indra.databases.chebi\_client, 92  
indra.databases.chembl\_client, 99  
indra.databases.context\_client, 94  
indra.databases.doid\_client, 105  
indra.databases.efo\_client, 104  
indra.databases.go\_client, 102  
indra.databases.hgnc\_client, 90  
indra.databases.hp\_client, 104  
indra.databases.lincs\_client, 101  
indra.databases.mesh\_client, 101  
indra.databases.mirbase\_client, 103  
indra.databases.ndex\_client, 95  
indra.databases.pubchem\_client, 103  
indra.databases.relevance\_client, 94  
indra.databases.uniprot\_client, 92  
indra.explanation.model\_checker.model\_checker, 161  
indra.explanation.model\_checker.pybel, 168  
indra.explanation.model\_checker.pysb, 165  
indra.explanation.model\_checker.signed\_graph, 167  
indra.explanation.model\_checker.unsigned\_graph, 168  
indra.literature, 105  
indra.literature.adeft\_tools, 112  
indra.literature.crossref\_client, 109  
indra.literature.elsevier\_client, 109  
indra.literature.newsapi\_client, 111  
indra.literature.pmc\_client, 108  
indra.literature.pubmed\_client, 106  
indra.mechlinker, 132  
indra.preassembler, 113  
indra.preassembler.grounding\_mapper, 121  
indra.preassembler.grounding\_mapper.adeft, 126  
indra.preassembler.grounding\_mapper.analysis, 124  
indra.preassembler.grounding\_mapper.mapper, 122  
indra.preassembler.hierarchy\_manager, 118  
indra.preassembler.ontology\_mapper, 129  
indra.preassembler.sitemapper, 126  
indra.sources.bel.api, 67  
indra.sources.bel.processor, 73  
indra.sources.bel.rdf\_processor, 69  
indra.sources.biogrid, 78  
indra.sources.biopax.api, 73  
indra.sources.biopax.pathway\_commons\_client, 76  
indra.sources.biopax.processor, 75  
indra.sources.cwms, 63

- indra.sources.cwms.api, 64
- indra.sources.cwms.processor, 64
- indra.sources.cwms.rdf\_processor, 65
- indra.sources.eidos, 59
- indra.sources.eidos.api, 60
- indra.sources.eidos.cli, 63
- indra.sources.eidos.processor, 61
- indra.sources.eidos.reader, 62
- indra.sources.eidos.server, 63
- indra.sources.geneways.api, 56
- indra.sources.geneways.processor, 57
- indra.sources.hprd, 78
- indra.sources.hprd.api, 79
- indra.sources.hprd.processor, 80
- indra.sources.hume, 67
- indra.sources.hume.api, 67
- indra.sources.hume.processor, 67
- indra.sources.indra\_db\_rest, 85
- indra.sources.indra\_db\_rest.api, 85
- indra.sources.indra\_db\_rest.processor, 88
- indra.sources.isi, 54
- indra.sources.isi.api, 54
- indra.sources.isi.processor, 56
- indra.sources.lincs\_drug, 83
- indra.sources.lincs\_drug.api, 83
- indra.sources.lincs\_drug.processor, 83
- indra.sources.medscan, 46
- indra.sources.medscan.api, 46
- indra.sources.medscan.processor, 47
- indra.sources.ndex\_cx.api, 83
- indra.sources.ndex\_cx.processor, 84
- indra.sources.reach.api, 36
- indra.sources.reach.processor, 38
- indra.sources.reach.reader, 39
- indra.sources.rlimsp, 58
- indra.sources.rlimsp.api, 58
- indra.sources.rlimsp.processor, 59
- indra.sources.signor.api, 77
- indra.sources.signor.processor, 78
- indra.sources.sofia, 66
- indra.sources.sofia.api, 66
- indra.sources.sofia.processor, 67
- indra.sources.sparser, 44
- indra.sources.sparser.api, 44
- indra.sources.tas, 82
- indra.sources.tas.api, 82
- indra.sources.tas.processor, 83
- indra.sources.tees.api, 51
- indra.sources.tees.processor, 52
- indra.sources.trips.api, 40
- indra.sources.trips.client, 42
- indra.sources.trips.drum\_reader, 43
- indra.sources.trips.processor, 41
- indra.sources.trrust, 81
- indra.sources.trrust.api, 81
- indra.sources.trrust.processor, 82
- indra.statements.statements, 13
- indra.tools.assemble\_corpus, 169
- indra.tools.executable\_subnetwork, 183
- indra.tools.gene\_network, 182
- indra.tools.incremental\_model, 184
- indra.tools.live\_curation, 180
- indra.tools.machine, 185
- indra.tools.reading.pmid\_reading.read\_pmids, 187
- indra.tools.reading.read\_files, 195
- indra.tools.reading.readers, 195
- indra.tools.reading.run\_drum\_reading, 197
- indra.tools.reading.submit\_reading\_pipeline, 198
- indra.tools.reading.util.export\_stmts\_by\_reach\_rule, 200
- indra.tools.reading.util.log\_analysis\_tools, 200
- indra.tools.reading.util.reporter, 200
- indra.tools.reading.util.script\_tools, 201
- indra.util.aws, 202
- indra.util.get\_version, 203
- indra.util.nested\_dict, 203
- indra.util.plot\_formatting, 204

## A

- Acetylation (class in *indra.statements.statements*), 21
- Activation (class in *indra.statements.statements*), 23
- ActiveForm (class in *indra.statements.statements*), 24
- activities\_by\_target() (in module *indra.databases.chembl\_client*), 99
- ActivityCondition (class in *indra.statements.statements*), 17
- add\_activity\_form() (*indra.assemblers.pysb.base\_agents.BaseAgent* method), 140
- add\_activity\_type() (*indra.assemblers.pysb.base\_agents.BaseAgent* method), 140
- add\_agent() (*indra.assemblers.kami.assembler.Nugget* method), 158
- add\_default\_initial\_conditions() (*indra.assemblers.pysb.assembler.PysbAssembler* method), 136
- add\_edge() (*indra.assemblers.kami.assembler.Nugget* method), 158
- add\_entry() (*indra.preassembler.hierarchy\_manager.YamlHierarchyManager* method), 121
- add\_image() (*indra.tools.reading.util.reporter.Reporter* method), 200
- add\_node() (*indra.assemblers.kami.assembler.Nugget* method), 158
- add\_result() (*indra.tools.reading.readers.Reader* method), 196
- add\_reverse\_effects() (*indra.assemblers.pysb.preassembler.PysbPreassembler* method), 140
- add\_rule\_to\_model() (in module *indra.assemblers.pysb.assembler*), 138
- add\_section() (*indra.tools.reading.util.reporter.Reporter* method), 200
- add\_site\_states() (*indra.assemblers.pysb.base\_agents.BaseAgent* method), 140
- add\_statements() (*indra.assemblers.cag.assembler.CAGAssembler* method), 149
- add\_statements() (*indra.assemblers.cx.assembler.CxAssembler* method), 141
- add\_statements() (*indra.assemblers.cyjs.assembler.CyJSAssembler* method), 148
- add\_statements() (*indra.assemblers.english.assembler.EnglishAssembler* method), 143
- add\_statements() (*indra.assemblers.graph.assembler.GraphAssembler* method), 144
- add\_statements() (*indra.assemblers.html.assembler.HtmlAssembler* method), 152
- add\_statements() (*indra.assemblers.index\_card.assembler.IndexCardAssembler* method), 146
- add\_statements() (*indra.assemblers.indranet.assembler.IndraNetAssembler* method), 160
- add\_statements() (*indra.assemblers.pysb.assembler.PysbAssembler* method), 136
- add\_statements() (*indra.assemblers.pysb.preassembler.PysbPreassembler* method), 140
- add\_statements() (*indra.assemblers.sbgm.assembler.SBGmAssembler* method), 147
- add\_statements() (*indra.explanation.model\_checker.model\_checker.ModelChecker* method), 162
- add\_statements() (*indra.mechlinker.MechLinker* method), 133

- add\_statements() (*indra.preassembler.Preassembler* method), 113  
 add\_statements() (*indra.tools.incremental\_model.IncrementalModel* method), 184  
 add\_text() (*indra.tools.reading.util.reporter.Reporter* method), 201  
 add\_typing() (*indra.assemblers.kami.assembler.Nugget* method), 158  
 AddModification (class in *indra.statements.statements*), 19  
 Agent (class in *indra.statements.statements*), 31  
 agent\_from\_entity() (*indra.sources.medscan.processor.MedscanProcessor* method), 48  
 agent\_list() (*indra.statements.statements.Association* method), 29  
 agent\_list() (*indra.statements.statements.Influence* method), 28  
 agent\_list() (*indra.statements.statements.Statement* method), 18  
 agent\_set (*indra.assemblers.pysb.assembler.PysbAssembler* attribute), 136  
 agent\_texts() (in module *indra.preassembler.grounding\_mapper.analysis*), 124  
 agent\_texts\_with\_grounding() (in module *indra.preassembler.grounding\_mapper.analysis*), 124  
 AgentState (class in *indra.mechlinker*), 132  
 align\_statements() (in module *indra.tools.assemble\_corpus*), 169  
 all\_agents() (in module *indra.preassembler.grounding\_mapper.analysis*), 125  
 all\_direct\_stmts (*indra.sources.bel.rdf\_processor.BelRdfProcessor* attribute), 70  
 all\_events (*indra.sources.reach.processor.ReachProcessor* attribute), 39  
 all\_indirect\_stmts (*indra.sources.bel.rdf\_processor.BelRdfProcessor* attribute), 70  
 analyze\_db\_reading() (in module *indra.tools.reading.util.log\_analysis\_tools*), 200  
 analyze\_reach\_log() (in module *indra.tools.reading.util.log\_analysis\_tools*), 200  
 api\_ruler (*indra.sources.reach.reader.ReachReader* attribute), 40  
 append\_warning() (*indra.assemblers.html.assembler.HtmlAssembler* method), 152  
 apply\_to() (*indra.mechlinker.AgentState* method), 132  
 assembled\_stmts (*indra.tools.incremental\_model.IncrementalModel* attribute), 184  
 Association (class in *indra.statements.statements*), 29  
 Autophosphorylation (class in *indra.statements.statements*), 20
- ## B
- BaseAgent (class in *indra.assemblers.pysb.base\_agents*), 140  
 BaseAgent (class in *indra.mechlinker*), 133  
 BaseAgentSet (class in *indra.assemblers.pysb.base\_agents*), 140  
 BaseAgentSet (class in *indra.mechlinker*), 133  
 basename (*indra.tools.gene\_network.GeneNetwork* attribute), 182  
 BatchReadingError, 198  
 BayesianScorer (class in *indra.belief*), 129  
 BeliefEngine (class in *indra.belief*), 129  
 BeliefPackage (class in *indra.belief*), 130  
 BeliefScorer (class in *indra.belief*), 130  
 BelRdfProcessor (class in *indra.sources.bel.rdf\_processor*), 69  
 BioContext (class in *indra.statements.statements*), 33  
 BiogridProcessor (class in *indra.sources.biogrid*), 78  
 BiopaxProcessor (class in *indra.sources.biopax.processor*), 75  
 BMIModel (class in *indra.assemblers.pysb.bmi\_wrapper*), 153  
 bound\_conditions (*indra.mechlinker.AgentState* attribute), 132  
 BoundCondition (class in *indra.statements.statements*), 16  
 build\_transitive\_closure() (*indra.preassembler.hierarchy\_manager.HierarchyManager* method), 118  
 build\_transitive\_closures() (*indra.preassembler.hierarchy\_manager.HierarchyManager* method), 119
- ## C
- CAG (*indra.assemblers.cag.assembler.CAGAssembler* attribute), 149  
 CAGAssembler (class in *indra.assemblers.cag.assembler*), 149  
 ch\_end (*indra.sources.medscan.processor.MedscanEntity* attribute), 47  
 ch\_start (*indra.sources.medscan.processor.MedscanEntity* attribute), 47



- change\_format() (in-dra.tools.reading.readers.Content method), 195
- change\_id() (indra.tools.reading.readers.Content method), 195
- check\_entitlement() (in module in-dra.literature.elsevier\_client), 109
- check\_grounding\_map() (in-dra.preassembler.grounding\_mapper.mapper.GroundingMapper static method), 122
- check\_model() (in-dra.explanation.model\_checker.model\_checker.ModelChecker method), 162
- check\_prior\_probs() (indra.belief.BeliefsScorer method), 130
- check\_prior\_probs() (indra.belief.SimpleScorer method), 131
- check\_statement() (in-dra.explanation.model\_checker.model\_checker.ModelChecker method), 162
- citation (indra.sources.reach.processor.ReachProcessor attribute), 38
- clear\_input() (in-dra.tools.reading.readers.ReachReader method), 196
- cm\_json\_to\_graph() (in module in-dra.assemblers.pysb.kappa\_util), 141
- combine\_duplicate\_stmts() (in-dra.preassembler.Preassembler method), 113
- combine\_duplicates() (in-dra.preassembler.Preassembler method), 114
- combine\_related() (in-dra.preassembler.Preassembler method), 114
- Complex (class in indra.statements.statements), 26
- complex\_monomers\_default() (in module in-dra.assemblers.pysb.assembler), 138
- complex\_monomers\_one\_step() (in module in-dra.assemblers.pysb.assembler), 138
- Concept (class in indra.statements.statements), 30
- concept (indra.statements.statements.Event attribute), 29
- connected\_subgraph() (in-dra.sources.tees.processor.TEESProcessor method), 52
- Content (class in indra.tools.reading.readers), 195
- Context (class in indra.statements.statements), 34
- context (indra.statements.statements.Event attribute), 30
- Conversion (class in indra.statements.statements), 28
- convert\_unit() (in-dra.statements.statements.QuantitativeState static method), 32
- converted\_direct\_stmts (in-dra.sources.bel.rdf\_processor.BelRdfProcessor attribute), 69
- converted\_indirect\_stmts (in-dra.sources.bel.rdf\_processor.BelRdfProcessor attribute), 69
- copy\_default\_config() (in module in-dra.tools.machine), 187
- Corpus (class in indra.tools.live\_curation), 180
- create\_mod\_site() (in-dra.assemblers.pysb.base\_agents.BaseAgent method), 140
- create\_network() (in module in-dra.databases.ndex\_client), 95
- create\_site() (in-dra.assemblers.pysb.base\_agents.BaseAgent method), 140
- ModelCheckers (indra.tools.live\_curation.Corpus attribute), 180
- CWMSError, 64
- CWMSProcessor (class in in-dra.sources.cwms.processor), 64
- CWMSRDFProcessor (class in in-dra.sources.cwms.rdf\_processor), 65
- Cx (indra.assemblers.cx.assembler.CxAssembler attribute), 141
- CxAssembler (class in in-dra.assemblers.cx.assembler), 141
- CyJSAssembler (class in in-dra.assemblers.cyjs.assembler), 148
- ## D
- db\_rest\_url (indra.assemblers.html.assembler.HtmlAssembler attribute), 152
- Deacetylation (class in in-dra.statements.statements), 21
- DecreaseAmount (class in in-dra.statements.statements), 27
- Defarnesylation (class in in-dra.statements.statements), 22
- degenerate\_stmts (in-dra.sources.bel.rdf\_processor.BelRdfProcessor attribute), 70
- Degeranylgeranylation (class in in-dra.statements.statements), 22
- Deglycosylation (class in in-dra.statements.statements), 21
- Dehydroxylation (class in in-dra.statements.statements), 21
- delta (indra.statements.statements.Event attribute), 30
- Demethylation (class in in-dra.statements.statements), 23

- Demyristoylation (class in in-dra.statements.statements), 22
- Depalmitoylation (class in in-dra.statements.statements), 22
- Dephosphorylation (class in in-dra.statements.statements), 21
- Deribosylation (class in in-dra.statements.statements), 22
- Desumoylation (class in in-dra.statements.statements), 21
- determine\_reach\_subtype() (in module in-dra.sources.reach.processor), 39
- Deubiquitination (class in in-dra.statements.statements), 22
- df (indra.sources.trrust.processor.TrrustProcessor attribute), 82
- digraph\_from\_df() (in-dra.assemblers.indranet.net.IndraNet class method), 158
- directly\_or\_indirectly\_related() (in-dra.preassembler.hierarchy\_manager.HierarchyManager method), 119
- doc\_id (indra.sources.cwms.processor.CWMSProcessor attribute), 64
- doc\_id (indra.sources.trips.processor.TripsProcessor attribute), 41
- doi\_query() (in module in-dra.literature.crossref\_client), 109
- download\_article() (in module in-dra.literature.elsevier\_client), 109
- download\_article\_from\_ids() (in module in-dra.literature.elsevier\_client), 109
- download\_from\_search() (in module in-dra.literature.elsevier\_client), 109
- draw\_im() (indra.explanation.model\_checker.pysb.PysbModelChecker method), 165
- draw\_stmt\_graph() (in module in-dra.statements.statements), 35
- drum\_system (indra.sources.trips.drum\_reader.DrumReader attribute), 43
- DrumReader (class in in-dra.sources.trips.drum\_reader), 43
- dump\_logs() (in module in-dra.util.aws), 202
- dump\_statements() (in module in-dra.tools.assemble\_corpus), 169
- dump\_stmt\_strings() (in module in-dra.tools.assemble\_corpus), 169
- E**
- edge\_properties (in-dra.assemblers.graph.assembler.GraphAssembler attribute), 144
- eidos\_reader (indra.sources.eidos.reader.EidosReader attribute), 62
- EidosProcessor (class in in-dra.sources.eidos.processor), 61
- EidosReader (class in in-dra.sources.eidos.reader), 62
- eliminate\_exact\_duplicates() (in-dra.sources.biopax.processor.BiopaxProcessor method), 75
- EmptyReader (class in in-dra.tools.reading.readers), 196
- english\_join() (in module in-dra.assemblers.english.assembler), 143
- EnglishAssembler (class in in-dra.assemblers.english.assembler), 142
- entities (indra.sources.medscan.processor.MedscanRelation attribute), 50
- entity\_matches\_key() (in-dra.statements.statements.Agent method), 31
- ev\_totals (indra.assemblers.html.assembler.HTMLAssembler attribute), 152
- Event (class in in-dra.statements.statements), 29
- extract\_from\_event() (in-dra.sources.cwms.processor.CWMSProcessor method), 65
- Evidence (class in in-dra.statements.statements), 31
- evidence\_random\_noise\_prior() (in module in-dra.belief), 132
- evidences (indra.belief.BeliefPackage attribute), 130
- existing\_edges (in-dra.assemblers.graph.assembler.GraphAssembler attribute), 144
- existing\_nodes (in-dra.assemblers.graph.assembler.GraphAssembler attribute), 144
- expand\_families() (in module in-dra.tools.assemble\_corpus), 170
- expand\_pagination() (in module in-dra.literature.pubmed\_client), 106
- export\_dict() (indra.util.nested\_dict.NestedDict method), 204
- export\_into\_python() (in-dra.assemblers.pysb.bmi\_wrapper.BMIModel method), 153
- export\_model() (in-dra.assemblers.pysb.assembler.PysbAssembler method), 137
- export\_to\_cytoscapejs() (in-dra.assemblers.cag.assembler.CAGAssembler method), 149
- extend\_with() (in-dra.preassembler.hierarchy\_manager.HierarchyManager method), 119
- extra\_annotations (in-dra.sources.isi.processor.IsiProcessor attribute), 56

`extract_and_process()` (in module `indra.sources.eidos.cli`), 63  
`extract_causal_relations()` (in module `indra.sources.cwms.processor.CWMSProcessor` method), 65  
`extract_causal_relations()` (in module `indra.sources.eidos.processor.EidosProcessor` method), 61  
`extract_events()` (in module `indra.sources.cwms.processor.CWMSProcessor` method), 65  
`extract_from_directory()` (in module `indra.sources.eidos.cli`), 63  
`extract_output()` (in module `indra.sources.tees.api`), 52  
`extract_paragraphs()` (in module `indra.literature.elsevier_client`), 110  
`extract_paragraphs()` (in module `indra.literature.pmc_client`), 108  
`extract_statement_from_query_result()` (`indra.sources.cwms.rdf_processor.CWMSRDFProcessor` method), 65  
`extract_statements()` (in module `indra.sources.cwms.rdf_processor.CWMSRDFProcessor` method), 65  
`extract_statements()` (in module `indra.sources.rlimsp.processor.RlimspProcessor` method), 59  
`extract_statements()` (in module `indra.sources.trrust.processor.TrrustProcessor` method), 82  
`extract_text()` (in module `indra.literature.elsevier_client`), 110  
`extract_text()` (in module `indra.literature.pmc_client`), 108  
`extracted_events` (in module `indra.sources.trips.processor.TripsProcessor` attribute), 41  
`extractions` (`indra.sources.trips.drum_reader.DrumReader` attribute), 43

**F**

`Farnesylation` (class in module `indra.statements.statements`), 22  
`filename` (in module `indra.sources.medscan.api`), 47  
`filter_belief()` (in module `indra.tools.assemble_corpus`), 170  
`filter_by_db_refs()` (in module `indra.tools.assemble_corpus`), 170  
`filter_by_type()` (in module `indra.tools.assemble_corpus`), 171  
`filter_concept_names()` (in module `indra.tools.assemble_corpus`), 171  
`filter_direct()` (in module `indra.tools.assemble_corpus`), 171  
`filter_enzyme_kinase()` (in module `indra.tools.assemble_corpus`), 171  
`filter_evidence_source()` (in module `indra.tools.assemble_corpus`), 172  
`filter_gene_list()` (in module `indra.tools.assemble_corpus`), 172  
`filter_genes_only()` (in module `indra.tools.assemble_corpus`), 172  
`filter_grounded_only()` (in module `indra.tools.assemble_corpus`), 173  
`filter_human_only()` (in module `indra.tools.assemble_corpus`), 173  
`filter_inconsequential_acts()` (in module `indra.tools.assemble_corpus`), 173  
`filter_inconsequential_mods()` (in module `indra.tools.assemble_corpus`), 174  
`filter_mod_nokinase()` (in module `indra.tools.assemble_corpus`), 174  
`filter_mutation_status()` (in module `indra.tools.assemble_corpus`), 174  
`filter_no_hypothesis()` (in module `indra.tools.assemble_corpus`), 175  
`filter_no_negated()` (in module `indra.tools.assemble_corpus`), 175  
`filter_paragraphs()` (in module `indra.literature.adeft_tools`), 112  
`filter_pmids()` (in module `indra.literature.pmc_client`), 108  
`filter_top_level()` (in module `indra.tools.assemble_corpus`), 175  
`filter_transcription_factor()` (in module `indra.tools.assemble_corpus`), 175  
`filter_uuid_list()` (in module `indra.tools.assemble_corpus`), 176  
`finalize()` (`indra.assemblers.pysb.bmi_wrapper.BMIModel` method), 153  
`find_arg()` (in module `indra.sources.eidos.processor`), 62  
`find_args()` (in module `indra.sources.eidos.processor`), 62  
`find_contradicts()` (in module `indra.preassembler.Preassembler` method), 116  
`find_entity` (`indra.preassembler.hierarchy_manager.HierarchyManager` attribute), 119  
`find_event_parent_with_event_child()` (`indra.sources.tees.processor.TEESProcessor` method), 52  
`find_event_with_outgoing_edges()` (in module `indra.sources.tees.processor.TEESProcessor` method), 53  
`find_paths()` (`indra.explanation.model_checker.model_checker.ModelChecker` method), 153

- method), 162
- `flatten_evidence()` (in module `indra.preassembler`), 116
- `flatten_stmts()` (in module `indra.preassembler`), 117
- `format_axis()` (in module `indra.util.plot_formatting`), 204
- `from_df()` (`indra.assemblers.indranet.net.IndraNet` class method), 158
- `from_file()` (`indra.tools.reading.readers.Content` class method), 195
- `from_seconds()` (`indra.statements.statements.QuantitativeState` static method), 32
- `from_string()` (`indra.tools.reading.readers.Content` class method), 195
- ## G
- `g` (`indra.sources.bel.rdf_processor.BelRdfProcessor` attribute), 69
- `Gap` (class in `indra.statements.statements`), 25
- `gather_explicit_activities()` (`indra.mechlinker.MechLinker` method), 133
- `gather_implicit_activities()` (`indra.mechlinker.MechLinker` method), 133
- `Gef` (class in `indra.statements.statements`), 25
- `gene_list` (`indra.tools.gene_network.GeneNetwork` attribute), 182
- `GeneNetwork` (class in `indra.tools.gene_network`), 182
- `general_node_label()` (`indra.sources.tees.processor.TEESProcessor` method), 53
- `generate_im()` (`indra.explanation.model_checker.pysb.PysbModelChecker` method), 165
- `generate_jupyter_js()` (`indra.assemblers.cag.assembler.CAGAssembler` method), 149
- `geneways_action_to_indra_statement_type` (in module `indra.sources.geneways.processor`), 57
- `GenewaysProcessor` (class in `indra.sources.geneways.processor`), 57
- `geo_context_from_ref()` (`indra.sources.eidos.processor.EidosProcessor` method), 61
- `Geranylgeranylation` (class in `indra.statements.statements`), 22
- `get()` (`indra.util.nested_dict.NestedDict` method), 204
- `get_abstract()` (in module `indra.literature.elsevier_client`), 110
- `get_abstract()` (in module `indra.literature.pubmed_client`), 106
- `get_activating_mods()` (`indra.sources.bel.rdf_processor.BelRdfProcessor` method), 70
- `get_activating_subs()` (`indra.sources.bel.rdf_processor.BelRdfProcessor` method), 70
- `get_activation()` (`indra.sources.bel.rdf_processor.BelRdfProcessor` method), 70
- `get_activation()` (`indra.sources.reach.processor.ReachProcessor` method), 39
- `get_activations()` (`indra.sources.trips.processor.TripsProcessor` method), 41
- `get_activations_causal()` (`indra.sources.trips.processor.TripsProcessor` method), 41
- `get_activations_stimulate()` (`indra.sources.trips.processor.TripsProcessor` method), 41
- `get_active_forms()` (`indra.sources.trips.processor.TripsProcessor` method), 41
- `get_active_forms_state()` (`indra.sources.trips.processor.TripsProcessor` method), 41
- `get_activity_modification()` (`indra.sources.biopax.processor.BiopaxProcessor` method), 75
- `get_ag_ns_id()` (in module `indra.assemblers.indranet.assembler`), 161
- `get_agent_from_entity_info()` (in module `indra.sources.rlimsp.processor`), 59
- `get_agent_rule_str()` (in module `indra.assemblers.pysb.assembler`), 138
- `get_agents()` (`indra.sources.ndex_cx.processor.NdexCxProcessor` method), 84
- `get_agents()` (`indra.sources.trips.processor.TripsProcessor` method), 41
- `get_agents_with_name()` (in module `indra.preassembler.grounding_mapper.analysis`), 125
- `get_all_descendants()` (in module `indra.statements.statements`), 35
- `get_all_direct_statements()` (`indra.sources.bel.rdf_processor.BelRdfProcessor` method), 71
- `get_all_events()` (`indra.sources.reach.processor.ReachProcessor` method), 39
- `get_all_events()` (`indra.sources.trips.processor.TripsProcessor` method), 42

`get_all_indirect_statements()` (in *dra.sources.bel.rdf\_processor.BelRdfProcessor* method), 71  
`get_annotation()` (in module *dra.assemblers.pysb.assembler*), 138  
`get_api_ruler()` (in *dra.sources.reach.reader.ReachReader* method), 40  
`get_article()` (in module *dra.literature.elsevier\_client*), 110  
`get_article_xml` (in module *dra.literature.pubmed\_client*), 106  
`get_attribute()` (in *dra.assemblers.pysb.bmi\_wrapper.BMIModel* method), 154  
`get_batch_command()` (in module *indra.util.aws*), 202  
`get_bel_stmts()` (in *dra.tools.gene\_network.GeneNetwork* method), 182  
`get_bio_hierarchies()` (in module *indra.preassembler.hierarchy\_manager*), 121  
`get_biopax_stmts()` (in *dra.tools.gene\_network.GeneNetwork* method), 182  
`get_cancer_studies()` (in module *indra.databases.cbio\_client*), 96  
`get_cancer_types()` (in module *indra.databases.cbio\_client*), 96  
`get_case_lists()` (in module *indra.databases.cbio\_client*), 96  
`get_causal_edge()` (in module *dra.assemblers.pybel.assembler*), 157  
`get_ccle_cna()` (in module *indra.databases.cbio\_client*), 96  
`get_ccle_lines_for_mutation()` (in module *indra.databases.cbio\_client*), 97  
`get_ccle_mrna()` (in module *indra.databases.cbio\_client*), 97  
`get_ccle_mutations()` (in module *indra.databases.cbio\_client*), 97  
`get_chebi_entry_from_web` (in module *indra.databases.chebi\_client*), 92  
`get_chebi_id_from_cas()` (in module *indra.databases.chebi\_client*), 92  
`get_chebi_id_from_hmdb()` (in module *indra.databases.chebi\_client*), 92  
`get_chebi_id_from_name()` (in module *indra.databases.chebi\_client*), 92  
`get_chebi_id_from_pubchem()` (in module *indra.databases.chebi\_client*), 92  
`get_chebi_name_from_id()` (in module *indra.databases.chebi\_client*), 92  
`get_chebi_name_from_id_web()` (in module *indra.databases.chebi\_client*), 92  
`get_chembl_id()` (in module *indra.databases.chembl\_client*), 99  
`get_children()` (in *dra.preassembler.hierarchy\_manager.HierarchyManager* method), 119  
`get_complexes()` (in *dra.sources.bel.rdf\_processor.BelRdfProcessor* method), 71  
`get_complexes()` (in *dra.sources.biopax.processor.BiopaxProcessor* method), 75  
`get_complexes()` (in *dra.sources.hprd.processor.HprdProcessor* method), 81  
`get_complexes()` (in *dra.sources.reach.processor.ReachProcessor* method), 39  
`get_complexes()` (in *dra.sources.trips.processor.TripsProcessor* method), 42  
`get_concept()` (in *dra.sources.eidos.processor.EidosProcessor* method), 61  
`get_conversions()` (in *dra.sources.bel.rdf\_processor.BelRdfProcessor* method), 71  
`get_conversions()` (in *dra.sources.biopax.processor.BiopaxProcessor* method), 75  
`get_corpus()` (*indra.tools.live\_curation.LiveCurator* method), 181  
`get_create_base_agent()` (in *dra.assemblers.pysb.base\_agents.BaseAgentSet* method), 140  
`get_create_base_agent()` (in *dra.mechlinker.BaseAgentSet* method), 133  
`get_create_parameter()` (in module *dra.assemblers.pysb.assembler*), 138  
`get_current_hgnc_id()` (in module *indra.databases.hgnc\_client*), 90  
`get_current_time()` (in *dra.assemblers.pysb.bmi\_wrapper.BMIModel* method), 154  
`get_default_ndex_cred()` (in module *indra.databases.ndex\_client*), 95  
`get_degenerate_statements()` (in *dra.sources.bel.rdf\_processor.BelRdfProcessor* method), 71  
`get_degradations()` (in *dra.sources.trips.processor.TripsProcessor* method), 42

`get_doid_id_from_doid_alt_id()` (in module `indra.databases.doid_client`), 105  
`get_doid_id_from_doid_name()` (in module `indra.databases.doid_client`), 105  
`get_doid_name_from_doid_id()` (in module `indra.databases.doid_client`), 105  
`get_dois` (in module `indra.literature.elsevier_client`), 110  
`get_drug_inhibition_stmts()` (in module `indra.databases.chembl_client`), 99  
`get_drug_target_data()` (in module `indra.databases.lincs_client`), 101  
`get_ecs_cluster_for_queue()` (in module `indra.tools.reading.submit_reading_pipeline`), 198  
`get_efo_id_from_efo_name()` (in module `indra.databases.efo_client`), 104  
`get_efo_name_from_efo_id()` (in module `indra.databases.efo_client`), 104  
`get_entity_text_for_relation()` (in module `indra.sources.tees.processor.TEESProcessor` method), 53  
`get_entrez_id()` (in module `indra.databases.hgnc_client`), 90  
`get_ev_count()` (in module `indra.sources.indra_db_rest.processor.IndraDBRestProcessor` method), 89  
`get_ev_count_by_hash()` (in module `indra.sources.indra_db_rest.processor.IndraDBRestProcessor` method), 89  
`get_ev_counts()` (in module `indra.sources.indra_db_rest.processor.IndraDBRestProcessor` method), 89  
`get_evidence()` (in module `indra.databases.chembl_client`), 99  
`get_evidence()` (in module `indra.sources.eidos.processor.EidosProcessor` method), 61  
`get_filename()` (in module `indra.tools.reading.readers.Content` method), 195  
`get_filepath()` (in module `indra.tools.reading.readers.Content` method), 196  
`get_full_text()` (in module `indra.literature`), 105  
`get_fulltext_links()` (in module `indra.literature.crossref_client`), 109  
`get_gap()` (in module `indra.sources.biopax.processor.BiopaxProcessor` method), 76  
`get_gef()` (in module `indra.sources.biopax.processor.BiopaxProcessor` method), 76  
`get_gene_names()` (in module `indra.assemblers.cyjs.assembler.CyJSAssembler` method), 148  
`get_genetic_profiles()` (in module `indra.databases.cbio_client`), 97  
`get_git_info()` (in module `indra.util.get_version`), 203  
`get_go_id_from_label()` (in module `indra.databases.go_client`), 102  
`get_go_label()` (in module `indra.databases.go_client`), 102  
`get_graph()` (`indra.explanation.model_checker.model_checker.ModelChecker` method), 162  
`get_graph()` (`indra.explanation.model_checker.pybel.PybelModelChecker` method), 169  
`get_graph()` (`indra.explanation.model_checker.pysb.PysbModelChecker` method), 165  
`get_graph()` (`indra.explanation.model_checker.signed_graph.SignedGraph` method), 167  
`get_graph()` (`indra.explanation.model_checker.unsigned_graph.UnsignedGraph` method), 168  
`get_grounded_agent()` (in module `indra.sources.trrust.processor`), 82  
`get_grounding()` (in module `indra.statements.statements.Agent` method), 31  
`get_groundings()` (in module `indra.sources.eidos.processor.EidosProcessor` method), 61  
`get_hash()` (`indra.statements.statements.Statement` method), 18  
`get_hash_statements_dict()` (in module `indra.sources.indra_db_rest.processor.IndraDBRestProcessor` method), 89  
`get_kernel()` (in module `indra.databases.relevance_client`), 94  
`get_hedging()` (in module `indra.sources.eidos.processor.EidosProcessor` static method), 62  
`get_hgnc_entry` (in module `indra.databases.hgnc_client`), 90  
`get_hgnc_from_entrez()` (in module `indra.databases.hgnc_client`), 90  
`get_hgnc_from_mouse()` (in module `indra.databases.hgnc_client`), 90  
`get_hgnc_from_rat()` (in module `indra.databases.hgnc_client`), 90  
`get_hgnc_id()` (in module `indra.databases.hgnc_client`), 90  
`get_hgnc_id_from_mirbase_id()` (in module `indra.databases.mirbase_client`), 103  
`get_hgnc_name()` (in module `indra.databases.hgnc_client`), 91  
`get_hp_id_from_hp_name()` (in module `indra.databases.hp_client`), 104  
`get_hp_name_from_hp_id()` (in module `indra.databases.hp_client`), 104

`get_id_count()` (in module `indra.literature.pubmed_client`), 106  
`get_ids` (in module `indra.literature.pubmed_client`), 106  
`get_ids_for_gene` (in module `indra.literature.pubmed_client`), 106  
`get_ids_for_mesh()` (in module `indra.literature.pubmed_client`), 107  
`get_im()` (`indra.explanation.model_checker.pysb.PysbModelChecker` method), 165  
`get_inchi_key` (in module `indra.databases.pubchem_client`), 103  
`get_inchi_key()` (in module `indra.databases.chebi_client`), 93  
`get_input_var_names()` (`indra.assemblers.pysb.bmi_wrapper.BMIModel` method), 154  
`get_issns_for_journal` (in module `indra.literature.pubmed_client`), 107  
`get_job_log()` (in module `indra.util.aws`), 202  
`get_jobs()` (in module `indra.util.aws`), 202  
`get_kinetics()` (in module `indra.databases.chembl_client`), 99  
`get_leaves()` (`indra.util.nested_dict.NestedDict` method), 204  
`get_log_by_name()` (in module `indra.util.aws`), 202  
`get_logs_from_db_reading()` (in module `indra.tools.reading.util.log_analysis_tools`), 200  
`get_mesh_id()` (in module `indra.databases.chembl_client`), 100  
`get_mesh_id_name()` (in module `indra.databases.mesh_client`), 101  
`get_mesh_id_name_from_web()` (in module `indra.databases.mesh_client`), 102  
`get_mesh_name()` (in module `indra.databases.mesh_client`), 102  
`get_mesh_name_from_web` (in module `indra.databases.mesh_client`), 102  
`get_metadata` (in module `indra.literature.crossref_client`), 109  
`get_metadata_for_ids()` (in module `indra.literature.pubmed_client`), 107  
`get_metadata_from_xml_tree()` (in module `indra.literature.pubmed_client`), 107  
`get_mirbase_id_from_hgnc_id()` (in module `indra.databases.mirbase_client`), 103  
`get_mirbase_id_from_hgnc_symbol()` (in module `indra.databases.mirbase_client`), 104  
`get_mirbase_id_from_mirbase_name()` (in module `indra.databases.mirbase_client`), 103  
`get_mirbase_name_from_mirbase_id()` (in module `indra.databases.mirbase_client`), 103  
`get_model_agents()` (`indra.tools.incremental_model.IncrementalModel` method), 184  
`get_modifications()` (`indra.sources.bel.rdf_processor.BelRdfProcessor` method), 71  
`get_modifications()` (`indra.sources.biopax.processor.BiopaxProcessor` method), 76  
`get_modifications()` (`indra.sources.reach.processor.ReachProcessor` method), 39  
`get_modifications()` (`indra.sources.trips.processor.TripsProcessor` method), 42  
`get_modifications_indirect()` (`indra.sources.trips.processor.TripsProcessor` method), 42  
`get_monomer_pattern()` (in module `indra.assemblers.pysb.assembler`), 139  
`get_mouse_id()` (in module `indra.databases.hgnc_client`), 91  
`get_mutations()` (in module `indra.databases.cbio_client`), 98  
`get_mutations()` (in module `indra.databases.context_client`), 94  
`get_negation()` (`indra.sources.eidos.processor.EidosProcessor` static method), 62  
`get_node_names()` (`indra.sources.ndex_cx.processor.NdexCxProcessor` method), 84  
`get_nugget_dict()` (`indra.assemblers.kami.assembler.Nugget` method), 158  
`get_num_sequenced()` (in module `indra.databases.cbio_client`), 98  
`get_output()` (`indra.tools.reading.readers.ReachReader` method), 196  
`get_output()` (`indra.tools.reading.readers.SparsesReader` method), 197  
`get_output_var_names()` (`indra.assemblers.pysb.bmi_wrapper.BMIModel` method), 154  
`get_parents()` (`indra.preassembler.hierarchy_manager.HierarchyManager` method), 119  
`get_parser()` (in module `indra.tools.reading.util.script_tools`), 201  
`get_path()` (`indra.util.nested_dict.NestedDict` method), 204  
`get_paths()` (`indra.util.nested_dict.NestedDict` method), 204  
`get_pcid()` (in module `indra.databases.chembl_client`), 100  
`get_piis()` (in module `indra`)

*dra.literature.elsevier\_client*), 110  
*get\_piis\_for\_date* (in module *indra.literature.elsevier\_client*), 111  
*get\_pmid*() (in module *indra.databases.chembl\_client*), 100  
*get\_pmids*() (*indra.sources.ndex\_cx.processor.NdexCxProcessor* method), 84  
*get\_ppis*() (*indra.sources.hprd.processor.HprdProcessor* method), 81  
*get\_profile\_data*() (in module *indra.databases.cbio\_client*), 98  
*get\_protein\_expression*() (in module *indra.databases.context\_client*), 94  
*get\_protein\_refs*() (*indra.databases.lincs\_client.LincsClient* method), 101  
*get\_protein\_targets\_only*() (in module *indra.databases.chembl\_client*), 100  
*get\_ptms*() (*indra.sources.hprd.processor.HprdProcessor* method), 81  
*get\_pubchem\_id*() (in module *indra.databases.chebi\_client*), 93  
*get\_publications*() (in module *indra.databases.biogrid\_client*), 93  
*get\_rat\_id*() (in module *indra.databases.hgnc\_client*), 91  
*get\_reader*() (in module *indra.tools.reading.readers*), 197  
*get\_reader\_class*() (in module *indra.tools.reading.readers*), 197  
*get\_reader\_classes*() (in module *indra.tools.reading.readers*), 197  
*get\_regulate\_activities*() (*indra.sources.biopax.processor.BiopaxProcessor* method), 76  
*get\_regulate\_amounts*() (*indra.sources.biopax.processor.BiopaxProcessor* method), 76  
*get\_regulate\_amounts*() (*indra.sources.reach.processor.ReachProcessor* method), 39  
*get\_regulate\_amounts*() (*indra.sources.trips.processor.TripsProcessor* method), 42  
*get\_related\_node*() (*indra.sources.tees.processor.TEESProcessor* method), 53  
*get\_relevant\_nodes*() (in module *indra.databases.relevance\_client*), 94  
*get\_s3\_file\_tree*() (in module *indra.util.aws*), 202  
*get\_sentences\_for\_agent*() (in module *indra.preassembler.grounding\_mapper.analysis*), 125  
*get\_site\_pattern*() (in module *indra.assemblers.pysb.assembler*), 139  
*get\_sites*() (*indra.sources.medscan.processor.ProteinSiteInfo* method), 50  
*get\_small\_molecule\_name*() (*indra.databases.lincs\_client.LincsClient* method), 101  
*get\_small\_molecule\_refs*() (*indra.databases.lincs\_client.LincsClient* method), 101  
*get\_source*() (*indra.assemblers.html.assembler.IndraHTMLLoader* method), 152  
*get\_source\_count*() (*indra.sources.indra\_db\_rest.processor.IndraDBRestProcessor* method), 89  
*get\_source\_count\_by\_hash*() (*indra.sources.indra\_db\_rest.processor.IndraDBRestProcessor* method), 89  
*get\_source\_counts*() (*indra.sources.indra\_db\_rest.processor.IndraDBRestProcessor* method), 89  
*get\_source\_hash*() (*indra.statements.statements.Evidence* method), 32  
*get\_specific\_id*() (in module *indra.databases.chebi\_client*), 93  
*get\_start\_time*() (*indra.assemblers.pysb.bmi\_wrapper.BMIModel* method), 154  
*get\_statement\_by\_name*() (in module *indra.statements.statements*), 35  
*get\_statements*() (in module *indra.sources.indra\_db\_rest.api*), 85  
*get\_statements*() (*indra.sources.isi.processor.IsiProcessor* method), 56  
*get\_statements*() (*indra.sources.ndex\_cx.processor.NdexCxProcessor* method), 85  
*get\_statements*() (*indra.tools.gene\_network.GeneNetwork* method), 183  
*get\_statements*() (*indra.tools.incremental\_model.IncrementalModel* method), 184  
*get\_statements*() (*indra.tools.reading.readers.ReadingData* method), 197  
*get\_statements\_by\_hash*() (in module *indra.sources.indra\_db\_rest.api*), 87  
*get\_statements\_for\_paper*() (in module *indra.sources.indra\_db\_rest.api*), 86  
*get\_statements\_noprior*() (*indra.tools.incremental\_model.IncrementalModel*



method), 185

get\_statements\_prior() (in *indra.tools.incremental\_model.IncrementalModel* method), 185

get\_status() (*indra.assemblers.pysb.bmi\_wrapper.BMIModel* method), 154

get\_stmts() (in *module indra.tools.reading.pmid\_reading.read\_pmid*s), 187

get\_string() (*indra.assemblers.graph.assembler.GraphAssemble*method), 155

get\_subnetwork() (in *module indra.tools.executable\_subnetwork*), 183

get\_syntheses() (in *indra.sources.trips.processor.TripsProcessor* method), 42

get\_target\_chemblid() (in *module indra.databases.chembl\_client*), 100

get\_term\_agents() (in *indra.sources.trips.processor.TripsProcessor* method), 42

get\_text() (*indra.tools.reading.readers.Content* method), 196

get\_text\_content\_for\_gene() (in *module indra.literature.adeft\_tools*), 112

get\_text\_content\_for\_pmid() (in *module indra.literature.adeft\_tools*), 112

get\_time\_step() (in *indra.assemblers.pysb.bmi\_wrapper.BMIModel* method), 154

get\_time\_units() (in *indra.assemblers.pysb.bmi\_wrapper.BMIModel* method), 154

get\_title() (in *module indra.literature.pubmed\_client*), 108

get\_transcription() (in *indra.sources.bel.rdf\_processor.BelRdfProcessor* method), 72

get\_translocation() (in *indra.sources.reach.processor.ReachProcessor* method), 39

get\_typing\_dict() (in *indra.assemblers.kami.assembler.Nugget* method), 158

get\_uncond\_agent() (in *module indra.assemblers.pysb.assembler*), 139

get\_uniprot\_id() (in *module indra.databases.hgnc\_client*), 91

get\_unresolved\_support\_uuids() (in *module indra.statements.statements*), 34

get\_unyielding\_tcids() (in *module indra.tools.reading.util.log\_analysis\_tools*), 200

get\_valid\_location() (in *module indra.statements.statements*), 35

get\_valid\_residue() (in *module indra.statements.statements*), 35

get\_value() (*indra.assemblers.pysb.bmi\_wrapper.BMIModel* method), 154

get\_values() (*indra.assemblers.pysb.bmi\_wrapper.BMIModel* method), 154

get\_var\_name() (in *indra.assemblers.pysb.bmi\_wrapper.BMIModel* method), 155

get\_var\_rank() (in *indra.assemblers.pysb.bmi\_wrapper.BMIModel* method), 155

get\_var\_type() (in *indra.assemblers.pysb.bmi\_wrapper.BMIModel* method), 155

get\_var\_units() (in *indra.assemblers.pysb.bmi\_wrapper.BMIModel* method), 155

get\_version() (in *module indra.sources.sparses.api*), 46

get\_version() (in *module indra.util.get\_version*), 203

get\_wm\_hierarchies() (in *module indra.preassembler.hierarchy\_manager*), 121

get\_xml() (in *module indra.literature.pmc\_client*), 108

get\_xml() (in *module indra.sources.trips.client*), 42

GetReadingStatsError, 200

gets() (*indra.util.nested\_dict.NestedDict* method), 204

Glycosylation (class in *indra.statements.statements*), 21

graph (*indra.assemblers.graph.assembler.GraphAssemble* attribute), 144

graph (*indra.assemblers.sif.assembler.SifAssemble* attribute), 145

graph (*indra.explanation.model\_checker.model\_checker.ModelChecker* attribute), 161

graph (*indra.preassembler.hierarchy\_manager.HierarchyManager* attribute), 118

graph\_properties (in *indra.assemblers.graph.assembler.GraphAssemble* attribute), 144

graph\_query() (in *module indra.sources.biopax.pathway\_commons\_client*), 76

GraphAssembler (class in *indra.assemblers.graph.assembler*), 144

grounded\_monomer\_patterns() (in *module indra.assemblers.pysb.assembler*), 139

GroundingMapper (class in *indra.preassembler.grounding\_mapper.mapper*), 122

GtpActivation (class in *indra*)

*dra.statements.statements*), 24

## H

*HasActivity* (class in *indra.statements.statements*), 25

*hierarchies* (*indra.preassembler.Preassembler* attribute), 113

*HierarchyManager* (class in *indra.preassembler.hierarchy\_manager*), 118

*HprdProcessor* (class in *indra.sources.hprd.processor*), 80

*HtmlAssembler* (class in *indra.assemblers.html.assembler*), 151

*HumeJsonLdProcessor* (class in *indra.sources.hume.processor*), 67

*Hydroxylation* (class in *indra.statements.statements*), 21

## I

*id\_df* (*indra.sources.hprd.processor.HprdProcessor* attribute), 80

*id\_lookup()* (in module *indra.literature*), 105

*id\_lookup()* (in module *indra.literature.pmc\_client*), 108

*im\_json\_to\_graph()* (in module *indra.assemblers.pysb.kappa\_util*), 141

*IncreaseAmount* (class in *indra.statements.statements*), 28

*IncrementalModel* (class in *indra.tools.incremental\_model*), 184

*IndexCardAssembler* (class in *indra.assemblers.index\_card.assembler*), 146

*indirect\_stmts* (*indra.sources.bel.rdf\_processor.BelRdfProcessor* attribute), 69

*indra.assemblers.cag.assembler* (module), 149

*indra.assemblers.cx.assembler* (module), 141

*indra.assemblers.cyjs.assembler* (module), 148

*indra.assemblers.english.assembler* (module), 142

*indra.assemblers.graph.assembler* (module), 144

*indra.assemblers.html.assembler* (module), 151

*indra.assemblers.index\_card.assembler* (module), 146

*indra.assemblers.indranet* (module), 158

*indra.assemblers.indranet.assembler* (module), 160

*indra.assemblers.indranet.net* (module), 158

*indra.assemblers.kami.assembler* (module), 157

*indra.assemblers.pybel.assembler* (module), 156

*indra.assemblers.pysb.assembler* (module), 135

*indra.assemblers.pysb.base\_agents* (module), 140

*indra.assemblers.pysb.bmi\_wrapper* (module), 153

*indra.assemblers.pysb.kappa\_util* (module), 141

*indra.assemblers.pysb.preassembler* (module), 140

*indra.assemblers.sbgm.assembler* (module), 147

*indra.assemblers.sif.assembler* (module), 145

*indra.assemblers.tsv.assembler* (module), 150

*indra.belief* (module), 129

*indra.databases.biogrid\_client* (module), 93

*indra.databases.cbio\_client* (module), 96

*indra.databases.chebi\_client* (module), 92

*indra.databases.chembl\_client* (module), 99

*indra.databases.context\_client* (module), 94

*indra.databases.doid\_client* (module), 105

*indra.databases.efo\_client* (module), 104

*indra.databases.go\_client* (module), 102

*indra.databases.hgnc\_client* (module), 90

*indra.databases.hp\_client* (module), 104

*indra.databases.lincs\_client* (module), 101

*indra.databases.mesh\_client* (module), 101

*indra.databases.mirbase\_client* (module), 103

*indra.databases.ndex\_client* (module), 95

*indra.databases.pubchem\_client* (module), 103

*indra.databases.relevance\_client* (module), 94

*indra.databases.uniprot\_client* (module), 92

*indra.explanation.model\_checker.model\_checker* (module), 161

*indra.explanation.model\_checker.pybel* (module), 168

*indra.explanation.model\_checker.pysb* (module), 165

*indra.explanation.model\_checker.signed\_graph* (module), 167

*indra.explanation.model\_checker.unsigned\_graph* (module), 168

- indra.literature (module), 105
- indra.literature.adeft\_tools (module), 112
- indra.literature.crossref\_client (module), 109
- indra.literature.elsevier\_client (module), 109
- indra.literature.newsapi\_client (module), 111
- indra.literature.pmc\_client (module), 108
- indra.literature.pubmed\_client (module), 106
- indra.mechlinker (module), 132
- indra.preassembler (module), 113
- indra.preassembler.grounding\_mapper (module), 121
- indra.preassembler.grounding\_mapper.adeft\_mapper (module), 126
- indra.preassembler.grounding\_mapper.analysis\_mapper (module), 124
- indra.preassembler.grounding\_mapper.mapping\_mapper (module), 122
- indra.preassembler.hierarchy\_manager (module), 118
- indra.preassembler.ontology\_mapper (module), 129
- indra.preassembler.sitemapper (module), 126
- indra.sources.bel.api (module), 67
- indra.sources.bel.processor (module), 73
- indra.sources.bel.rdf\_processor (module), 69
- indra.sources.biogrid (module), 78
- indra.sources.biopax.api (module), 73
- indra.sources.biopax.pathway\_commons\_client (module), 76
- indra.sources.biopax.processor (module), 75
- indra.sources.cwms (module), 63
- indra.sources.cwms.api (module), 64
- indra.sources.cwms.processor (module), 64
- indra.sources.cwms.rdf\_processor (module), 65
- indra.sources.eidos (module), 59
- indra.sources.eidos.api (module), 60
- indra.sources.eidos.cli (module), 63
- indra.sources.eidos.processor (module), 61
- indra.sources.eidos.reader (module), 62
- indra.sources.eidos.server (module), 63
- indra.sources.geneways.api (module), 56
- indra.sources.geneways.processor (module), 57
- indra.sources.hprd (module), 78
- indra.sources.hprd.api (module), 79
- indra.sources.hprd.processor (module), 80
- indra.sources.hume (module), 67
- indra.sources.hume.api (module), 67
- indra.sources.hume.processor (module), 67
- indra.sources.indra\_db\_rest (module), 85
- indra.sources.indra\_db\_rest.api (module), 85
- indra.sources.indra\_db\_rest.processor (module), 88
- indra.sources.isi (module), 54
- indra.sources.isi.api (module), 54
- indra.sources.isi.processor (module), 56
- indra.sources.lincs\_drug (module), 83
- indra.sources.lincs\_drug.api (module), 83
- indra.sources.lincs\_drug.processor (module), 83
- indra.sources.medscan (module), 46
- indra.sources.medscan.api (module), 46
- indra.sources.medscan.processor (module), 47
- indra.sources.ndex\_cx.api (module), 83
- indra.sources.ndex\_cx.processor (module), 84
- indra.sources.reach.api (module), 36
- indra.sources.reach.processor (module), 38
- indra.sources.reach.reader (module), 39
- indra.sources.rlimsp (module), 58
- indra.sources.rlimsp.api (module), 58
- indra.sources.rlimsp.processor (module), 59
- indra.sources.signor.api (module), 77
- indra.sources.signor.processor (module), 78
- indra.sources.sofia (module), 66
- indra.sources.sofia.api (module), 66
- indra.sources.sofia.processor (module), 67
- indra.sources.sparser (module), 44
- indra.sources.sparser.api (module), 44
- indra.sources.tas (module), 82
- indra.sources.tas.api (module), 82
- indra.sources.tas.processor (module), 83
- indra.sources.tees.api (module), 51
- indra.sources.tees.processor (module), 52
- indra.sources.trips.api (module), 40
- indra.sources.trips.client (module), 42
- indra.sources.trips.drum\_reader (module), 43
- indra.sources.trips.processor (module), 41
- indra.sources.trrust (module), 81
- indra.sources.trrust.api (module), 81
- indra.sources.trrust.processor (module), 82
- indra.statements.statements (module), 13
- indra.tools.assemble\_corpus (module), 169

- `indra.tools.executable_subnetwork (module)`, 183
- `indra.tools.gene_network (module)`, 182
- `indra.tools.incremental_model (module)`, 184
- `indra.tools.live_curation (module)`, 180
- `indra.tools.machine (module)`, 185
- `indra.tools.reading.pmid_reading.read_pmid (module)`, 187
- `indra.tools.reading.read_files (module)`, 195
- `indra.tools.reading.readers (module)`, 195
- `indra.tools.reading.run_drug_reading (module)`, 197
- `indra.tools.reading.submit_reading_pipeline (module)`, 198
- `indra.tools.reading.util.export_stmts_by_reach (module)`, 200
- `indra.tools.reading.util.log_analysis_tools (module)`, 200
- `indra.tools.reading.util.reporter (module)`, 200
- `indra.tools.reading.util.script_tools (module)`, 201
- `indra.util.aws (module)`, 202
- `indra.util.get_version (module)`, 203
- `indra.util.nested_dict (module)`, 203
- `indra.util.plot_formatting (module)`, 204
- `IndraDBRestProcessor (class in indra.sources.indra_db_rest.processor)`, 88
- `IndraHTMLLoader (class in indra.assemblers.html.assembler)`, 152
- `IndraNet (class in indra.assemblers.indranet.net)`, 158
- `IndraNetAssembler (class in indra.assemblers.indranet.assembler)`, 160
- `infer_activations () (indra.mechlinker.MechLinker static method)`, 134
- `infer_active_forms () (indra.mechlinker.MechLinker static method)`, 134
- `infer_complexes () (indra.mechlinker.MechLinker static method)`, 134
- `infer_modifications () (indra.mechlinker.MechLinker static method)`, 134
- `Influence (class in indra.statements.statements)`, 28
- `influence_from_event () (indra.sources.cwms.processor.CWMSProcessor method)`, 65
- `influence_from_relation () (indra.sources.cwms.processor.CWMSProcessor method)`, 65
- `Inhibition (class in indra.statements.statements)`, 23
- `initialize () (indra.assemblers.pysb.bmi_wrapper.BMIModel method)`, 155
- `initialize_reader () (indra.sources.eidos.api)`, 60
- `initialize_reader () (indra.sources.eidos.reader.EidosReader method)`, 62
- `InputError`, 30
- `interval (in module indra.sources.medscan.api)`, 47
- `invalid_site_pos (indra.sources.hprd.processor.HprdProcessor attribute)`, 81
- `InvalidCorpusError`, 181
- `InvalidLocationError`, 30
- `InvalidResidueError`, 30
- `is_format () (indra.tools.reading.readers.Content method)`, 196
- `is_kinase () (indra.databases.hgnc_client)`, 91
- `is_opposite () (indra.preassembler.hierarchy_manager.HierarchyManager method)`, 119
- `is_phosphatase () (indra.databases.hgnc_client)`, 91
- `is_transcription_factor () (indra.databases.hgnc_client)`, 91
- `is_working () (indra.sources.indra_db_rest.processor.IndraDBRestProcessor method)`, 89
- `isa () (indra.preassembler.hierarchy_manager.HierarchyManager method)`, 120
- `isa_or_partof () (indra.preassembler.hierarchy_manager.HierarchyManager method)`, 120
- `IsiProcessor (class in indra.sources.isi.processor)`, 56
- `items () (indra.assemblers.pysb.base_agents.BaseAgentSet method)`, 140
- ## J
- `join_json_files () (indra.tools.reading.pmid_reading.read_pmid)`, 187
- ## K
- `KamiAssembler (class in indra.assemblers.kami.assembler)`, 157
- `kill_all () (indra.util.aws)`, 203
- ## L
- `last_site_info_in_sentence (indra.sources.medscan.processor.MedscanProcessor attribute)`, 48
- `lazy (in module indra.sources.medscan.api)`, 47

length (*indra.explanation.model\_checker.model\_checker.PathResult* attribute), 163

LincsClient (class in *indra.databases.lincs\_client*), 101

LincsProcessor (class in *indra.sources.lincs\_drug.processor*), 83

LinkedStatement (class in *indra.mechlinker*), 133

LiveCurator (class in *indra.tools.live\_curation*), 181

load\_from\_rdf\_file() (in *indra.preassembler.hierarchy\_manager.HierarchyManager* method), 120

load\_from\_rdf\_graph() (in *indra.preassembler.hierarchy\_manager.HierarchyManager* method), 120

load\_from\_rdf\_string() (in *indra.preassembler.hierarchy\_manager.HierarchyManager* method), 120

load\_go\_graph() (in module *indra.databases.go\_client*), 102

load\_grounding\_map() (in module *indra.preassembler.grounding\_mapper.mapper*), 124

load\_lincs\_csv() (in module *indra.databases.lincs\_client*), 101

load\_prior() (*indra.tools.incremental\_model.IncrementalModel* method), 185

load\_statements() (in module *indra.tools.assemble\_corpus*), 176

location (*indra.mechlinker.AgentState* attribute), 132

**M**

make\_df() (*indra.assemblers.indranet.assembler.IndraNetAssembler* method), 160

make\_generic\_copy() (in *indra.statements.statements.Statement* method), 18

make\_hash() (in module *indra.statements.statements*), 35

make\_model() (*indra.assemblers.cag.assembler.CAGAssembler* method), 149

make\_model() (*indra.assemblers.cx.assembler.CxAssembler* method), 141

make\_model() (*indra.assemblers.cyjs.assembler.CyJSAssembler* method), 148

make\_model() (*indra.assemblers.english.assembler.EnglishAssembler* method), 143

make\_model() (*indra.assemblers.graph.assembler.GraphAssembler* method), 145

make\_model() (*indra.assemblers.html.assembler.HtmlAssembler* method), 152

make\_model() (*indra.assemblers.index\_card.assembler.IndexCardAssembler* method), 146

make\_model() (*indra.assemblers.indranet.assembler.IndraNetAssembler* method), 160

make\_model() (*indra.assemblers.kami.assembler.KamiAssembler* method), 157

make\_model() (*indra.assemblers.pysb.assembler.PysbAssembler* method), 137

make\_model() (*indra.assemblers.sbgm.assembler.SBGNAssembler* method), 147

make\_model() (*indra.assemblers.sif.assembler.SifAssembler* method), 145

make\_model() (*indra.assemblers.tsv.assembler.TsvAssembler* method), 151

make\_nxml\_from\_text() (in module *indra.sources.sparser.api*), 46

make\_parser() (in module *indra.tools.reading.read\_files*), 195

make\_report() (in *indra.tools.reading.util.reporter.Reporter* method), 201

make\_repository\_component() (in *indra.assemblers.pysb.bmi\_wrapper.BMIModel* method), 155

make\_statement() (in *indra.sources.geneways.processor.GenewaysProcessor* method), 57

make\_statement\_camel() (in module *indra.statements.statements*), 35

make\_stmt() (in module *indra.sources.trrust.processor*), 82

many\_ups\_for\_refseq (in *indra.sources.hprd.processor.HprdProcessor* attribute), 81

map\_agent() (*indra.preassembler.grounding\_mapper.mapper.GroundingMapper* method), 122

map\_agents\_for\_stmt() (in *indra.preassembler.grounding\_mapper.mapper.GroundingMapper* method), 122

map\_grounding() (in module *indra.tools.assemble\_corpus*), 176

map\_sequence() (in module *indra.tools.assemble\_corpus*), 176

map\_sites() (*indra.preassembler.sitemapper.SiteMapper* method), 128

map\_statements() (in *indra.preassembler.ontology\_mapper.OntologyMapper* method), 129

map\_stmts() (*indra.preassembler.grounding\_mapper.mapper.GroundingMapper* method), 123

MappedStatement (class in *indra.preassembler.sitemapper*), 126

matches\_key() (*indra.statements.statements.AgentState* method), 31

max\_data\_length (in *indra.explanation.model\_checker.model\_checker.PathResult* attribute), 164

max\_paths (*indra.explanation.model\_checker.model\_checker.PathResult* attribute), 164

- attribute*), 164
  - MechLinker (*class in indra.mechlinker*), 133
  - MedscanEntity (*class in indra.sources.medscan.processor*), 47
  - MedscanProcessor (*class in indra.sources.medscan.processor*), 48
  - MedscanProperty (*class in indra.sources.medscan.processor*), 49
  - MedscanRelation (*class in indra.sources.medscan.processor*), 50
  - merge\_deltas() (*in module indra.tools.assemble\_corpus*), 177
  - merge\_groundings() (*in module indra.tools.assemble\_corpus*), 177
  - merge\_results() (*in indra.sources.indra\_db\_rest.processor.IndraDBRestProcessor* *method*), 89
  - metadata (*indra.assemblers.html.assembler.HtmlAssembler* *attribute*), 152
  - Methylation (*class in indra.statements.statements*), 23
  - Migration (*class in indra.statements.statements*), 30
  - migration\_from\_event() (*in indra.sources.cwms.processor.CWMSProcessor* *method*), 65
  - ModCondition (*class in indra.statements.statements*), 16
  - model (*indra.assemblers.english.assembler.EnglishAssembler* *attribute*), 143
  - model (*indra.assemblers.html.assembler.HtmlAssembler* *attribute*), 152
  - model (*indra.assemblers.indranet.assembler.IndraNetAssembler* *attribute*), 160
  - model (*indra.assemblers.pysb.assembler.PysbAssembler* *attribute*), 136
  - model (*indra.sources.biopax.processor.BiopaxProcessor* *attribute*), 75
  - model\_to\_owl() (*in module indra.sources.biopax.pathway\_commons\_client*), 77
  - ModelChecker (*class in indra.explanation.model\_checker.model\_checker*), 161
  - Modification (*class in indra.statements.statements*), 19
  - mods (*indra.mechlinker.AgentState* *attribute*), 132
  - motif\_window (*indra.sources.hprd.processor.HprdProcessor* *attribute*), 81
  - MovementContext (*class in indra.statements.statements*), 34
  - mutations (*indra.mechlinker.AgentState* *attribute*), 132
  - MutCondition (*class in indra.statements.statements*), 16
  - Myristoylation (*class in indra.statements.statements*), 22
- ## N
- name (*indra.assemblers.pysb.assembler.Param* *attribute*), 135
  - name (*indra.assemblers.pysb.assembler.Policy* *attribute*), 136
  - name (*indra.sources.medscan.processor.MedscanEntity* *attribute*), 48
  - name (*indra.sources.medscan.processor.MedscanProperty* *attribute*), 49
  - namespace\_from\_uri() (*in module indra.sources.bel.rdf\_processor*), 72
  - NdexCxProcessor (*class in indra.sources.ndex\_cx.processor*), 84
  - NestedDict (*class in indra.util.nested\_dict*), 203
  - network\_name (*indra.assemblers.cx.assembler.CxAssembler* *attribute*), 141
  - no\_hgnc\_for\_egid (*in indra.sources.hprd.processor.HprdProcessor* *attribute*), 80
  - no\_mech\_ctr (*indra.sources.signor.processor.SignorProcessor* *attribute*), 78
  - no\_mech\_rows (*indra.sources.signor.processor.SignorProcessor* *attribute*), 78
  - no\_up\_for\_hgnc (*in indra.sources.hprd.processor.HprdProcessor* *attribute*), 80
  - no\_up\_for\_refseq (*in indra.sources.hprd.processor.HprdProcessor* *attribute*), 80
  - node\_has\_edge\_with\_label() (*in indra.sources.tees.processor.TEESProcessor* *method*), 53
  - node\_properties (*in indra.assemblers.graph.assembler.GraphAssembler* *attribute*), 144
  - node\_to\_evidence() (*in indra.sources.tees.processor.TEESProcessor* *method*), 53
  - normalize\_medscan\_name() (*in module indra.sources.medscan.processor*), 50
  - NotAStatementName, 30
  - Nugget (*class in indra.assemblers.kami.assembler*), 158
  - num\_entities (*indra.sources.medscan.processor.MedscanProcessor* *attribute*), 48
  - num\_entities\_not\_found (*in indra.sources.medscan.processor.MedscanProcessor* *attribute*), 48
- ## O
- obj (*indra.sources.medscan.processor.MedscanRelation* *attribute*), 50

*off\_by\_one* (*indra.sources.hprd.processor.HprdProcessor* attribute), 81  
*OntologyMapper* (class in *indra.preassembler.ontology\_mapper*), 129  
*owl\_str\_to\_model()* (in module *indra.sources.biopax.pathway\_commons\_client*), 77  
*owl\_to\_model()* (in module *indra.sources.biopax.pathway\_commons\_client*), 77  
**P**  
*Palmitoylation* (class in *indra.statements.statements*), 22  
*par\_to\_sec* (*indra.sources.cwms.processor.CWMSProcessor* attribute), 65  
*par\_to\_sec* (*indra.sources.trips.processor.TripsProcessor* attribute), 41  
*paragraphs* (*indra.sources.cwms.processor.CWMSProcessor* attribute), 65  
*paragraphs* (*indra.sources.trips.processor.TripsProcessor* attribute), 41  
*Param* (class in *indra.assemblers.pysb.assembler*), 135  
*parameters* (*indra.assemblers.pysb.assembler.Policy* attribute), 136  
*parse\_identifiers\_url()* (in module *indra.assemblers.pysb.assembler*), 139  
*partof()* (*indra.preassembler.hierarchy\_manager.HierarchyManager* method), 120  
*path\_found* (*indra.explanation.model\_checker.model\_checker.PathResult* attribute), 164  
*path\_metrics* (*indra.explanation.model\_checker.model\_checker.PathResult* attribute), 164  
*PathMetric* (class in *indra.explanation.model\_checker.model\_checker*), 163  
*PathResult* (class in *indra.explanation.model\_checker.model\_checker*), 164  
*paths* (*indra.explanation.model\_checker.model\_checker.PathResult* attribute), 164  
*Phosphorylation* (class in *indra.statements.statements*), 20  
*physical\_only* (in *indra.sources.biogrid.BiogridProcessor* attribute), 78  
*pmid* (*indra.sources.isi.processor.IsiProcessor* attribute), 56  
*pmid* (*indra.sources.medscan.processor.MedscanRelation* attribute), 50  
*policies* (*indra.assemblers.pysb.assembler.PysbAssembler* attribute), 136  
*Policy* (class in *indra.assemblers.pysb.assembler*), 136  
*position* (*indra.sources.reach.processor.Site* attribute), 39  
*preassemble()* (*indra.tools.incremental\_model.IncrementalModel* method), 185  
*Preassembler* (class in *indra.preassembler*), 113  
*prep\_input()* (*indra.tools.reading.readers.ReachReader* method), 196  
*prep\_input()* (*indra.tools.reading.readers.SparsesReader* method), 197  
*print\_boolean\_net()* (*indra.assemblers.sif.assembler.SifAssembler* method), 145  
*print\_cx()* (*indra.assemblers.cx.assembler.CxAssembler* method), 142  
*print\_cyjs\_context()* (*indra.assemblers.cyjs.assembler.CyJSAssembler* method), 148  
*print\_cyjs\_graph()* (*indra.assemblers.cyjs.assembler.CyJSAssembler* method), 148  
*print\_event\_statistics()* (*indra.sources.reach.processor.ReachProcessor* method), 39  
*print\_loopy()* (*indra.assemblers.sif.assembler.SifAssembler* method), 146  
*print\_model()* (*indra.assemblers.index\_card.assembler.IndexCardAssembler* method), 146  
*print\_model()* (*indra.assemblers.pysb.assembler.PysbAssembler* method), 137  
*print\_model()* (*indra.assemblers.sbg.assembler.SBGAssembler* method), 147  
*print\_model()* (*indra.assemblers.sif.assembler.SifAssembler* method), 146  
*print\_parent\_and\_children\_info()* (*indra.sources.tees.processor.TEESProcessor* method), 53  
*print\_statement\_coverage()* (*indra.sources.bel.rdf\_processor.BelRdfProcessor* method), 72  
*print\_statements()* (*indra.sources.bel.rdf\_processor.BelRdfProcessor* method), 72  
*print\_statements()* (*indra.sources.biopax.processor.BiopaxProcessor* method), 76  
*process\_belrdf()* (in module *indra.sources.bel.api*), 67  
*process\_belscript()* (in module *indra*), 67

*dra.sources.bel.api*), 68  
 process\_binding\_statements() (in-  
*dra.sources.tees.processor.TEESProcessor*  
 method), 53  
 process\_cbn\_jgif\_file() (in module in-  
*dra.sources.bel.api*), 68  
 process\_csv() (in module *indra.sources.tas.api*), 82  
 process\_csxml\_file() (in-  
*dra.sources.medscan.processor.MedscanProcessor*  
 method), 49  
 process\_cx() (in module *indra.sources.ndex\_cx.api*),  
 83  
 process\_cx\_file() (in module in-  
*dra.sources.ndex\_cx.api*), 83  
 process\_decrease\_expression\_amount()  
 (*indra.sources.tees.processor.TEESProcessor*  
 method), 53  
 process\_directory() (in module in-  
*dra.sources.medscan.api*), 46  
 process\_directory\_statements\_sorted\_by\_pmid()  
 (in module *indra.sources.medscan.api*), 47  
 process\_ekb() (in module *indra.sources.cwms.api*),  
 64  
 process\_ekb\_file() (in module in-  
*dra.sources.cwms.api*), 64  
 process\_file() (in module in-  
*dra.sources.medscan.api*), 47  
 process\_file\_sorted\_by\_pmid() (in module  
*indra.sources.medscan.api*), 47  
 process\_flat\_files() (in module in-  
*dra.sources.hprd.api*), 79  
 process\_from\_file() (in module in-  
*dra.sources.signor.api*), 77  
 process\_from\_json\_file() (in module in-  
*dra.sources.rlimsp.api*), 58  
 process\_from\_jsonish\_str() (in module in-  
*dra.sources.rlimsp.api*), 59  
 process\_from\_web() (in module in-  
*dra.sources.lincs\_drug.api*), 83  
 process\_from\_web() (in module in-  
*dra.sources.trrust.api*), 81  
 process\_from\_webservice() (in module in-  
*dra.sources.rlimsp.api*), 58  
 process\_geneways\_files() (in module in-  
*dra.sources.geneways.api*), 56  
 process\_increase\_expression\_amount()  
 (*indra.sources.tees.processor.TEESProcessor*  
 method), 54  
 process\_json() (in module *indra.sources.eidos.api*),  
 60  
 process\_json() (in module *indra.sources.sofia.api*),  
 66  
 process\_json\_dict() (in module in-  
*dra.sources.sparser.api*), 45  
 process\_json\_file() (in module in-  
*dra.sources.bel.api*), 68  
 process\_json\_file() (in module in-  
*dra.sources.eidos.api*), 60  
 process\_json\_file() (in module in-  
*dra.sources.isi.api*), 54  
 process\_json\_file() (in module in-  
*dra.sources.reach.api*), 36  
 process\_json\_file() (in module in-  
*dra.sources.sofia.api*), 66  
 process\_json\_str() (in module in-  
*dra.sources.eidos.api*), 60  
 process\_json\_str() (in module in-  
*dra.sources.reach.api*), 36  
 process\_jsonld() (in module in-  
*dra.sources.hume.api*), 67  
 process\_jsonld\_file() (in module in-  
*dra.sources.hume.api*), 67  
 process\_model() (in module in-  
*dra.sources.biopax.api*), 75  
 process\_ndex\_neighborhood() (in module in-  
*dra.sources.bel.api*), 68  
 process\_ndex\_network() (in module in-  
*dra.sources.ndex\_cx.api*), 84  
 process\_nxml() (in module *indra.sources.isi.api*), 54  
 process\_nxml\_file() (in module in-  
*dra.sources.reach.api*), 36  
 process\_nxml\_file() (in module in-  
*dra.sources.sparser.api*), 45  
 process\_nxml\_str() (in module in-  
*dra.sources.reach.api*), 37  
 process\_nxml\_str() (in module in-  
*dra.sources.sparser.api*), 44  
 process\_output\_folder() (in module in-  
*dra.sources.isi.api*), 55  
 process\_owl() (in module *indra.sources.biopax.api*),  
 74  
 process\_owl\_str() (in module in-  
*dra.sources.biopax.api*), 74  
 process\_pc\_neighborhood() (in module in-  
*dra.sources.biopax.api*), 73  
 process\_pc\_pathsbetween() (in module in-  
*dra.sources.biopax.api*), 73  
 process\_pc\_pathsfromto() (in module in-  
*dra.sources.biopax.api*), 74  
 process\_phosphorylation\_statements()  
 (*indra.sources.tees.processor.TEESProcessor*  
 method), 54  
 process\_pmc() (in module *indra.sources.reach.api*),  
 37  
 process\_preprocessed() (in module in-  
*dra.sources.isi.api*), 55  
 process\_pubmed\_abstract() (in module in-  
*dra.sources.reach.api*), 37



process\_pybel\_graph (in module in-dra.sources.bel.api), 69  
 process\_pybel\_neighborhood() (in module in-dra.sources.bel.api), 69  
 process\_rdf\_file() (in module in-dra.sources.cwms.api), 64  
 process\_relation() (in-dra.sources.medscan.processor.MedscanProcessor method), 49  
 process\_sparsener\_output() (in module in-dra.sources.sparsener.api), 45  
 process\_statement() (in-dra.explanation.model\_checker.model\_checker.ModelChecker method), 162  
 process\_statement() (in-dra.explanation.model\_checker.pybel.PybelModelChecker method), 169  
 process\_statement() (in-dra.explanation.model\_checker.pysb.PysbModelChecker method), 165  
 process\_statement() (in-dra.explanation.model\_checker.signed\_graph.SignedGraphModelChecker method), 167  
 process\_statement() (in-dra.explanation.model\_checker.unsigned\_graph.UnsignedGraphModelChecker method), 168  
 process\_subject() (in-dra.explanation.model\_checker.model\_checker.ModelChecker method), 163  
 process\_subject() (in-dra.explanation.model\_checker.pybel.PybelModelChecker method), 169  
 process\_subject() (in-dra.explanation.model\_checker.pysb.PysbModelChecker method), 166  
 process\_subject() (in-dra.explanation.model\_checker.signed\_graph.SignedGraphModelChecker method), 167  
 process\_subject() (in-dra.explanation.model\_checker.unsigned\_graph.UnsignedGraphModelChecker method), 168  
 process\_table() (in module in-dra.sources.sofia.api), 66  
 process\_text() (in module in-dra.sources.cwms.api), 64  
 process\_text() (in module indra.sources.eidos.api), 61  
 process\_text() (in module indra.sources.isi.api), 55  
 process\_text() (in module in-dra.sources.reach.api), 38  
 process\_text() (in module indra.sources.sofia.api), 66  
 process\_text() (in module in-dra.sources.sparsener.api), 44  
 process\_text() (in module indra.sources.tees.api), 51  
 process\_text() (in module indra.sources.trips.api), 40  
 process\_text() (in-dra.sources.eidos.reader.EidosReader method), 62  
 process\_xml() (in module in-dra.sources.sparsener.api), 45  
 process\_xml() (in module indra.sources.trips.api), 40  
 process\_xml\_file() (in module in-dra.sources.trips.api), 40  
 properties (indra.sources.medscan.processor.MedscanEntity attribute), 48  
 in\_map\_from\_twg() (in module in-dra.preassembler.grounding\_mapper.analysis), 125  
 InSiteInfo (class in in-dra.sources.medscan.processor), 50  
 prune\_influence\_map() (in-dra.explanation.model\_checker.pysb.PysbModelChecker method), 166  
 prune\_influence\_map\_degrade\_bind\_positive() (in-dra.explanation.model\_checker.pysb.PysbModelChecker method), 166  
 prune\_influence\_map\_subj\_obj() (in-dra.explanation.model\_checker.pysb.PysbModelChecker method), 166  
 prune\_nodes() (in-dra.explanation.model\_checker.model\_checker.ModelChecker method), 163  
 PybelAssembler (class in in-dra.assemblers.pybel.assembler), 156  
 PybelModelChecker (class in in-dra.explanation.model\_checker.pybel), 168  
 SignedGraphModelChecker (class in in-dra.sources.bel.processor), 73  
 PysbAssembler (class in in-dra.assemblers.pysb.assembler), 136  
 PysbModelChecker (class in in-dra.explanation.model\_checker.pysb), 165  
 PysbPreassembler (class in in-dra.assemblers.pysb.preassembler), 140

## Q

QualitativeDelta (class in in-dra.statements.statements), 32  
 QuantitativeState (class in in-dra.statements.statements), 32  
 query\_target() (in module in-dra.databases.chembl\_client), 100

## R

- ReachError, 196
- ReachOfflineReadingError, 39
- ReachProcessor (class in *indra.sources.reach.processor*), 38
- ReachReader (class in *indra.sources.reach.reader*), 39
- ReachReader (class in *indra.tools.reading.readers*), 196
- read() (*indra.tools.reading.readers.ReachReader* method), 196
- read() (*indra.tools.reading.readers.Reader* method), 196
- read() (*indra.tools.reading.readers.SparsedReader* method), 197
- read() (*indra.tools.reading.readers.TripsReader* method), 197
- read\_files() (in module *indra.tools.reading.read\_files*), 195
- read\_pmc() (*indra.sources.trips.drum\_reader.DrumReader* method), 44
- read\_pmid() (in module *indra.tools.reading.pmid\_reading.read\_pmid*), 187
- read\_pmid\_sentences() (in module *indra.tools.reading.run\_drum\_reading*), 197
- read\_some() (*indra.tools.reading.readers.SparsedReader* method), 197
- read\_text() (in module *indra.tools.reading.run\_drum\_reading*), 198
- read\_text() (*indra.sources.trips.drum\_reader.DrumReader* method), 44
- Reader (class in *indra.tools.reading.readers*), 196
- reader\_output (in *indra.sources.isi.processor.IsiProcessor* attribute), 56
- ReadingData (class in *indra.tools.reading.readers*), 196
- ReadingError, 197
- receive\_reply() (in *indra.sources.trips.drum\_reader.DrumReader* method), 44
- reduce\_activities() (in module *indra.tools.assemble\_corpus*), 177
- reduce\_activities() (in *indra.mechlinker.MechLinker* method), 135
- ref\_context\_from\_geoloc() (in module *indra.sources.eidos.processor*), 62
- RefContext (class in *indra.statements.statements*), 33
- RegulateActivity (class in *indra.statements.statements*), 23
- RegulateAmount (class in *indra.statements.statements*), 27
- related\_stmts (*indra.preassembler.Preassembler* attribute), 113
- remove\_im\_params() (in module *indra.explanation.model\_checker.pysb*), 166
- RemoveModification (class in *indra.statements.statements*), 19
- rename\_agents() (in *indra.preassembler.grounding\_mapper.mapper.GroundingMapper* static method), 123
- rename\_db\_ref() (in module *indra.tools.assemble\_corpus*), 178
- render\_stmt\_graph() (in module *indra.preassembler*), 117
- replace\_activations() (in *indra.mechlinker.MechLinker* method), 135
- replace\_activities() (in *indra.assemblers.pysb.preassembler.PysbPreassembler* method), 140
- replace\_complexes() (in *indra.mechlinker.MechLinker* method), 135
- Reporter (class in *indra.tools.reading.util.reporter*), 200
- require\_active\_forms() (in *indra.mechlinker.MechLinker* method), 135
- reset\_scorer() (in *indra.tools.live\_curation.LiveCurator* method), 181
- residue (*indra.sources.reach.processor.Site* attribute), 39
- result\_code (*indra.explanation.model\_checker.model\_checker.PathRes* attribute), 164
- results (*indra.tools.gene\_network.GeneNetwork* attribute), 182
- Ribosylation (class in *indra.statements.statements*), 22
- RlimspParagraph (class in *indra.sources.rlimsp.processor*), 59
- RlimspProcessor (class in *indra.sources.rlimsp.processor*), 59
- run\_adeft\_disambiguation() (in module *indra.preassembler.grounding\_mapper.adeft*), 126
- run\_eidos() (in module *indra.sources.eidos.cli*), 63
- run\_on\_text() (in module *indra.sources.tees.api*), 51
- run\_preassembly() (in module *indra.tools.assemble\_corpus*), 178
- run\_preassembly() (in *indra.tools.gene\_network.GeneNetwork* method), 183
- run\_preassembly\_duplicate() (in module *indra.tools.assemble\_corpus*), 179
- run\_preassembly\_related() (in module *indra.tools.assemble\_corpus*), 179
- run\_reach() (in module *indra.tools.reading.pmid\_reading.read\_pmid*), 187

run\_sparsifier() (in module *indra.sources.sparsifier.api*), 46  
 run\_sparsifier() (in module *indra.tools.reading.pmid\_reading.read\_pmid*), 188  
**S**  
 s2a() (in module *indra.sources.tees.processor*), 54  
 s3\_get() (*indra.tools.live\_curation.Corpora* method), 180  
 s3\_put() (*indra.tools.live\_curation.Corpora* method), 180  
 sample\_statements() (in module *indra.belief*), 132  
 save() (*indra.tools.incremental\_model.IncrementalModelSelfModification* (class in *indra.tools.incremental\_model*)), 185  
 save\_base\_map() (in module *indra.preassembler.grounding\_mapper.analysis*), 125  
 save\_dot() (*indra.assemblers.graph.assembler.GraphAssembler* method), 145  
 save\_json() (*indra.assemblers.cyjs.assembler.CyJSAssembler* method), 148  
 save\_model() (*indra.assemblers.cx.assembler.CxAssembler* method), 142  
 save\_model() (*indra.assemblers.cyjs.assembler.CyJSAssembler* method), 148  
 save\_model() (*indra.assemblers.html.assembler.HtmlAssembler* method), 152  
 save\_model() (*indra.assemblers.index\_card.assembler.IndexCardAssembler* method), 146  
 save\_model() (*indra.assemblers.pybel.assembler.PybelAssembler* method), 157  
 save\_model() (*indra.assemblers.pysb.assembler.PysbAssembler* method), 137  
 save\_model() (*indra.assemblers.sbgm.assembler.SBGmAssembler* method), 147  
 save\_model() (*indra.assemblers.sif.assembler.SifAssembler* method), 146  
 save\_model() (*indra.sources.biopax.processor.BiopaxProcessor* method), 76  
 save\_pdf() (*indra.assemblers.graph.assembler.GraphAssembler* method), 145  
 save\_rst() (*indra.assemblers.pysb.assembler.PysbAssembler* method), 138  
 save\_sentences() (in module *indra.preassembler.grounding\_mapper.analysis*), 125  
 save\_xml() (in module *indra.sources.trips.client*), 42  
 sbgm (*indra.assemblers.sbgm.assembler.SBGmAssembler* attribute), 147  
 SBGMAssembler (class in *indra.assemblers.sbgm.assembler*), 147  
 score\_evidence\_list() (*indra.belief.SimpleScorer* method), 131  
 score\_paths() (*indra.explanation.model\_checker.pysb.PysbModelChecker* method), 166  
 score\_statement() (*indra.belief.BeliefScorer* method), 130  
 score\_statement() (*indra.belief.SimpleScorer* method), 131  
 scorer (*indra.belief.BeliefEngine* attribute), 129  
 search\_science\_direct (in module *indra.literature.elsevier\_client*), 111  
 sec (*indra.sources.medscan.processor.MedscanRelation* attribute), 50  
 send\_query() (in module *indra.databases.chembl\_client*), 100  
 send\_query() (in module *indra.sources.trips.client*), 42  
 send\_request (in module *indra.databases.cbio\_client*), 99  
 send\_request() (in module *indra.databases.ndex\_client*), 95  
 send\_request() (in module *indra.literature.newsapi\_client*), 111  
 sentence\_statements (*indra.sources.medscan.processor.MedscanProcessor* attribute), 48  
 sentence\_statements (*indra.sources.cwms.processor.CWMSProcessor* attribute), 65  
 sentence\_statements (*indra.sources.trips.processor.TripsProcessor* attribute), 41  
 send\_reach\_logs() (in module *indra.tools.reading.util.log\_analysis\_tools*), 200  
 seq\_dict (*indra.sources.hprd.processor.HprdProcessor* attribute), 80  
 set\_base\_initial\_condition() (in module *indra.assemblers.pysb.assembler*), 139  
 set\_CCLE\_context() (*indra.assemblers.cyjs.assembler.CyJSAssembler* method), 149  
 set\_context() (*indra.assemblers.cx.assembler.CxAssembler* method), 142  
 set\_context() (*indra.assemblers.pysb.assembler.PysbAssembler* method), 138  
 set\_expression() (*indra.assemblers.pysb.assembler.PysbAssembler* method), 138  
 set\_extended\_initial\_condition() (in module *indra.assemblers.pysb.assembler*), 139

set\_fig\_params() (in module *indra.util.plot\_formatting*), 204  
 set\_hierarchy\_probs() (in *indra.belief.BeliefEngine* method), 130  
 set\_linked\_probs() (*indra.belief.BeliefEngine* method), 130  
 set\_location() (in *indra.tools.reading.readers.Content* method), 196  
 set\_prior\_probs() (*indra.belief.BeliefEngine* method), 130  
 set\_section\_order() (in *indra.tools.reading.util.reporter.Reporter* method), 201  
 set\_style() (in module *indra.databases.ndex\_client*), 95  
 set\_title() (*indra.tools.reading.util.reporter.Reporter* method), 201  
 set\_value() (*indra.assemblers.pysb.bmi\_wrapper.BMIModel* method), 155  
 set\_values() (*indra.assemblers.pysb.bmi\_wrapper.BMIModel* method), 155  
 SifAssembler (class in *indra.assemblers.sif.assembler*), 145  
 signed\_edges\_to\_signed\_nodes() (in *indra.explanation.model\_checker.model\_checker.ModelChecker* method), 163  
 signed\_from\_df() (in *indra.assemblers.indranet.net.IndraNet* class method), 159  
 SignedGraphModelChecker (class in *indra.explanation.model\_checker.signed\_graph*), 167  
 SignorProcessor (class in *indra.sources.signor.processor*), 78  
 SimpleScorer (class in *indra.belief*), 131  
 Site (class in *indra.sources.reach.processor*), 39  
 SiteMapper (class in *indra.preassembler.sitemapper*), 127  
 sites (*indra.assemblers.pysb.assembler.Policy* attribute), 136  
 source\_node (*indra.explanation.model\_checker.model\_checker.PatMIB* attribute), 163  
 SparserError, 197  
 SparserReader (class in *indra.tools.reading.readers*), 197  
 standardize\_agent\_name() (in *indra.preassembler.grounding\_mapper.mapper.GroundingMapper* static method), 123  
 standardize\_db\_refs() (in *indra.preassembler.grounding\_mapper.mapper.GroundingMapper* static method), 123  
 standardize\_names\_groundings() (in module *indra.tools.assemble\_corpus*), 180  
 state\_matches\_key() (in *indra.statements.statements.Agent* method), 31  
 Statement (class in *indra.statements.statements*), 17  
 statement\_base\_verb() (in module *indra.assemblers.english.assembler*), 143  
 statement\_key (*indra.belief.BeliefPackage* attribute), 130  
 statement\_passive\_verb() (in module *indra.assemblers.english.assembler*), 143  
 statement\_present\_verb() (in module *indra.assemblers.english.assembler*), 143  
 statements (*indra.assemblers.cag.assembler.CAGAssembler* attribute), 149  
 statements (*indra.assemblers.cx.assembler.CxAssembler* attribute), 141  
 statements (*indra.assemblers.cyjs.assembler.CyJSAAssembler* attribute), 148  
 statements (*indra.assemblers.english.assembler.EnglishAssembler* attribute), 143  
 statements (*indra.assemblers.graph.assembler.GraphAssembler* attribute), 144  
 statements (*indra.assemblers.html.assembler.HtmlAssembler* attribute), 152  
 statements (*indra.assemblers.index\_card.assembler.IndexCardAssembler* attribute), 146  
 statements (*indra.assemblers.pysb.assembler.PysbAssembler* attribute), 136  
 statements (*indra.assemblers.sbgm.assembler.SBGNAAssembler* attribute), 147  
 statements (*indra.assemblers.tsv.assembler.TsvAssembler* attribute), 151  
 statements (*indra.sources.bel.processor.PybelProcessor* attribute), 73  
 statements (*indra.sources.bel.rdf\_processor.BelRdfProcessor* attribute), 69  
 statements (*indra.sources.biogrid.BiogridProcessor* attribute), 78  
 statements (*indra.sources.biopax.processor.BiopaxProcessor* attribute), 75  
 statements (*indra.sources.cwms.processor.CWMSProcessor* attribute), 65  
 statements (*indra.sources.cwms.rdf\_processor.CWMSRDFProcessor* attribute), 65  
 statements (*indra.sources.eidos.processor.EidosProcessor* attribute), 61  
 statements (*indra.sources.geneways.processor.GenewaysProcessor* attribute), 57  
 statements (*indra.sources.hprd.processor.HprdProcessor* attribute), 80  
 statements (*indra.sources.hume.processor.HumeJsonLdProcessor* attribute), 67  
 statements (*indra.sources.indra\_db\_rest.processor.IndraDBRestProcessor* attribute), 89

statements (*indra.sources.isi.processor.IsiProcessor* attribute), 56  
 statements (*indra.sources.lincs\_drug.processor.LincsProcessor* attribute), 83  
 statements (*indra.sources.medscan.processor.MedscanProcessor* attribute), 48  
 statements (*indra.sources.ndex\_cx.processor.NdexCxProcessor* attribute), 84  
 statements (*indra.sources.reach.processor.ReachProcessor* attribute), 38  
 statements (*indra.sources.signor.processor.SignorProcessor* attribute), 78  
 statements (*indra.sources.tees.processor.TEESProcessor* attribute), 52  
 statements (*indra.sources.trips.processor.TripsProcessor* attribute), 41  
 statements (*indra.sources.trrust.processor.TrrustProcessor* attribute), 82  
 statements (*indra.tools.live\_curation.Corpus* attribute), 180  
 statements\_sample (*indra.sources.indra\_db\_rest.processor.IndraDBRestProcessor* attribute), 89  
 stmt\_type () (in module *indra.statements.statements*), 35  
 stmts (*indra.preassembler.Preassembler* attribute), 113  
 stmts (*indra.tools.incremental\_model.IncrementalModel* attribute), 184  
 stmts\_from\_json () (in module *indra.statements.statements*), 34  
 stmts\_from\_json\_file () (in module *indra.statements.statements*), 35  
 stmts\_to\_json () (in module *indra.statements.statements*), 34  
 stmts\_to\_json\_file () (in module *indra.statements.statements*), 35  
 strip\_agent\_context () (in module *indra.tools.assemble\_corpus*), 180  
 subj (*indra.sources.medscan.processor.MedscanRelation* attribute), 50  
 submit\_combine () (in module *indra.tools.reading.submit\_reading\_pipeline*), 198  
 submit\_curation () (in module *indra.sources.indra\_db\_rest.api*), 87  
 submit\_curation () (in module *indra.tools.live\_curation.LiveCurator* method), 181  
 submit\_reading () (in module *indra.tools.reading.submit\_reading\_pipeline*), 198  
 Sumoylation (class in *indra.statements.statements*), 21  
 svo\_type (*indra.sources.medscan.processor.MedscanRelation* attribute), 50  
**T**  
 tag\_evidence\_subtype () (in module *indra.belief*), 132  
 tag\_instance () (in module *indra.util.aws*), 203  
 tag\_instances\_on\_cluster () (in module *indra.tools.reading.submit\_reading\_pipeline*), 198  
 tag\_myself () (in module *indra.util.aws*), 203  
 tag\_text () (in module *indra.assemblers.html.assembler*), 153  
 tagged\_sentence (*indra.sources.medscan.processor.MedscanRelation* attribute), 50  
 target\_node (*indra.explanation.model\_checker.model\_checker.PathMethod* attribute), 163  
 TasProcessor (class in *indra.sources.tas.processor*), 83  
 TEESProcessor (class in *indra.sources.tees.processor*), 52  
 time\_context\_from\_uri () (in module *indra.sources.bel.rdf\_processor*), 73  
 time\_context\_from\_ref () (in module *indra.sources.eidos.processor.EidosProcessor* method), 62  
 time\_context\_from\_timex () (in module *indra.sources.eidos.processor*), 62  
 TimeContext (class in *indra.statements.statements*), 33  
 to\_database () (in module *indra.assemblers.pybel.assembler.PybelAssembler* method), 157  
 to\_digraph () (*indra.assemblers.indranet.net.IndraNet* method), 159  
 to\_graph () (*indra.statements.statements.Statement* method), 18  
 to\_impl () (in module *indra.sources.biopax.processor*), 76  
 to\_json () (*indra.statements.statements.ActiveForm* method), 24  
 to\_json () (*indra.statements.statements.Association* method), 29  
 to\_json () (*indra.statements.statements.Complex* method), 26  
 to\_json () (*indra.statements.statements.Conversion* method), 29  
 to\_json () (*indra.statements.statements.Event* method), 30  
 to\_json () (*indra.statements.statements.Evidence* method), 32  
 to\_json () (*indra.statements.statements.Gap* method), 26

- to\_json() (*indra.statements.statements.Gef* method), 25
- to\_json() (*indra.statements.statements.Influence* method), 28
- to\_json() (*indra.statements.statements.Modification* method), 19
- to\_json() (*indra.statements.statements.RegulateActivity* method), 23
- to\_json() (*indra.statements.statements.RegulateAmount* method), 27
- to\_json() (*indra.statements.statements.SelfModification* method), 20
- to\_json() (*indra.statements.statements.Statement* method), 18
- to\_json() (*indra.statements.statements.Translocation* method), 27
- to\_signed\_graph() (*indra.assemblers.indranet.net.IndraNet* method), 159
- to\_web() (*indra.assemblers.pybel.assembler.PybelAssembler* method), 157
- Translocation (class in *indra.statements.statements*), 27
- Transphosphorylation (class in *indra.statements.statements*), 20
- tree (*indra.sources.cwms.processor.CWMSProcessor* attribute), 64
- tree (*indra.sources.hume.processor.HumeJsonLdProcessor* attribute), 67
- tree (*indra.sources.reach.processor.ReachProcessor* attribute), 38
- tree (*indra.sources.trips.processor.TripsProcessor* attribute), 41
- TripsProcessor (class in *indra.sources.trips.processor*), 41
- TripsReader (class in *indra.tools.reading.readers*), 197
- TrrustProcessor (class in *indra.sources.trrust.processor*), 82
- TsvAssembler (class in *indra.assemblers.tsv.assembler*), 150
- type (*indra.sources.medscan.processor.MedscanEntity* attribute), 48
- type (*indra.sources.medscan.processor.MedscanProperty* attribute), 49
- U**
- Ubiquitination (class in *indra.statements.statements*), 22
- ungrounded\_texts() (in module *indra.preassembler.grounding\_mapper.analysis*), 126
- unique (*indra.assemblers.pysb.assembler.Param* attribute), 136
- unique\_stmts (*indra.preassembler.Preassembler* attribute), 113
- universal\_extract\_paragraphs() (in module *indra.literature.adeft\_tools*), 112
- universal\_extract\_text() (in module *indra.literature.adeft\_tools*), 112
- UnknownNamespaceException, 121
- UnknownPolicyError, 138
- Unresolved (class in *indra.statements.statements*), 29
- UnresolvedUuidError, 30
- UnsignedGraphModelChecker (class in *indra.explanation.model\_checker.unsigned\_graph*), 168
- update() (*indra.assemblers.pysb.bmi\_wrapper.BMIModel* method), 156
- update\_agent\_db\_refs() (*indra.preassembler.grounding\_mapper.mapper.GroundingMapper* method), 123
- update\_beliefs() (*indra.tools.live\_curation.LiveCurator* method), 181
- update\_counts() (*indra.belief.BayesianScorer* method), 129
- update\_id\_mappings() (in module *indra.databases.go\_client*), 103
- update\_network() (in module *indra.databases.ndex\_client*), 95
- update\_probs() (*indra.belief.BayesianScorer* method), 129
- update\_secondary\_mappings() (in module *indra.databases.go\_client*), 103
- upload\_model() (*indra.assemblers.cx.assembler.CxAssembler* method), 142
- urn (*indra.sources.medscan.processor.MedscanEntity* attribute), 48
- urn (*indra.sources.medscan.processor.MedscanProperty* attribute), 49
- V**
- value (*indra.assemblers.pysb.assembler.Param* attribute), 136
- value\_per\_second() (*indra.statements.statements.QuantitativeState* static method), 33
- verb (*indra.sources.medscan.processor.MedscanRelation* attribute), 50
- verbs (*indra.sources.isi.processor.IsiProcessor* attribute), 56
- W**
- wait\_for\_complete() (in module *indra.tools.reading.submit\_reading\_pipeline*), 199

`wait_until_done()` (*indra.sources.indra\_db\_rest.processor.IndraDBRestProcessor* method), 89

`WorldContext` (*class in indra.statements.statements*), 33

## Y

`YamlHierarchyManager` (*class in indra.preassembler.hierarchy\_manager*), 121