Ontologies tutorial:OPA2VEC

Fatima Zohra Smaili, Robert Hoehndorf

Semantic similarity (Limitations)

- Uses only a limited set of axioms from the ontology.
- Pre-defined measure for all datasets and applications.
- Reduces all information to one single value.

OPA2Vec: An Alternative …

- What is it ?
- How does it work ?
- How to use it ?

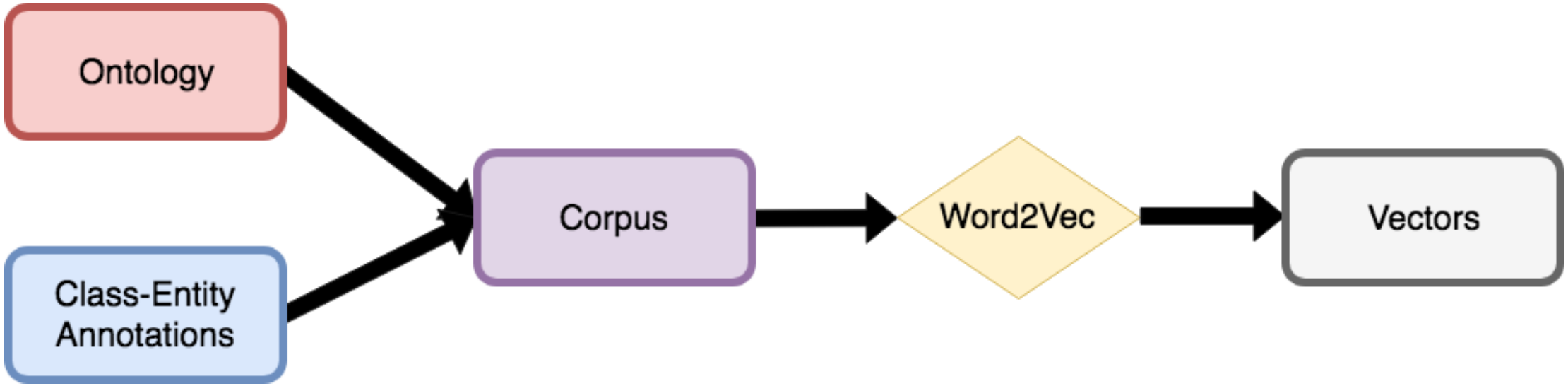
OPA2Vec

- Produces vector representations for biological entities from:
  - Ontologies: axioms + metadata (GO)
  - Their annotations to biological entities (proteins).

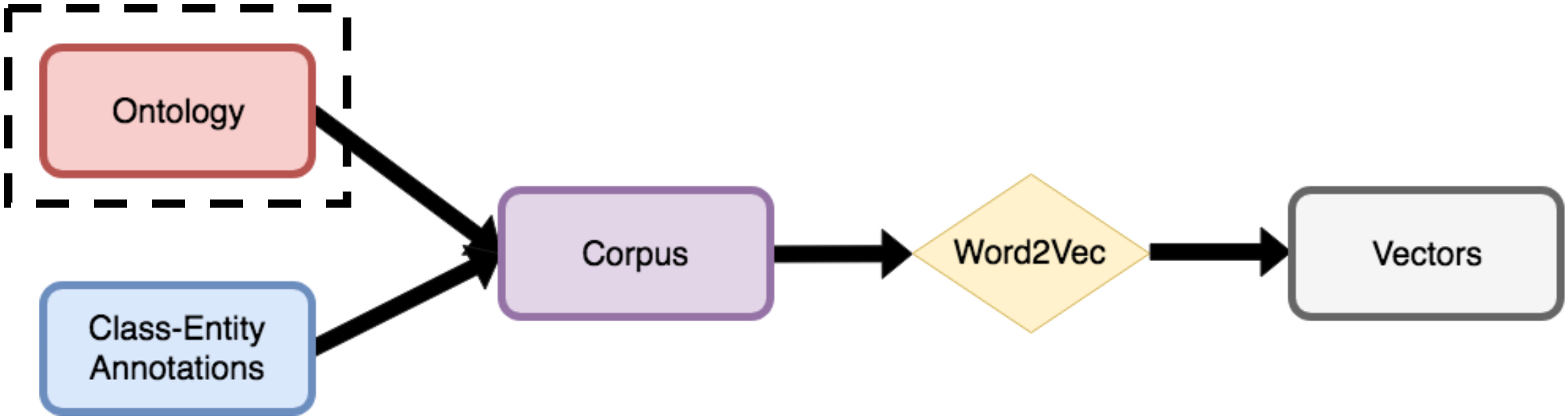
What can it be used for ?

- A **trained** and **data-specific** semantic similarity measure.
- Set of features for machine learning algorithms.
- Can be used to visualize the data.

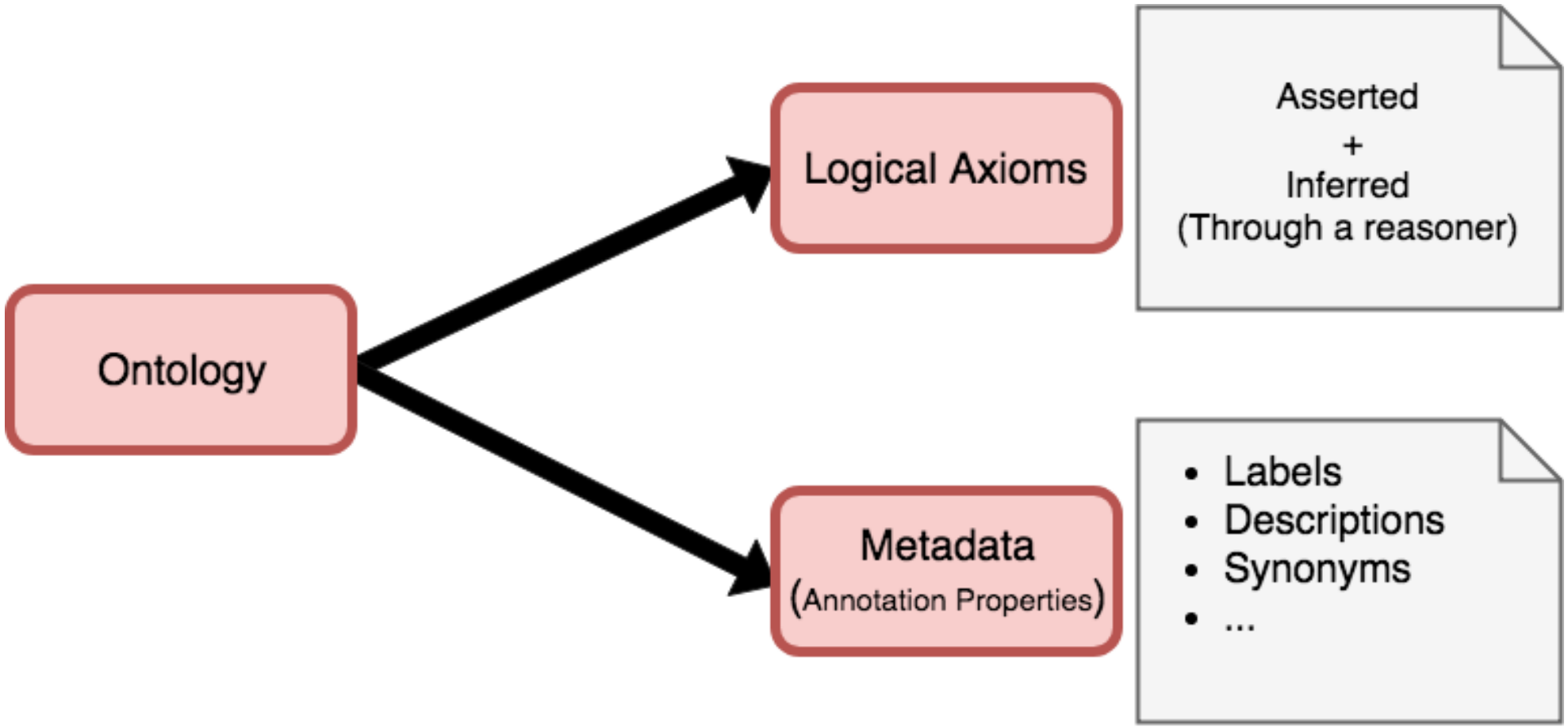
How does it work ?



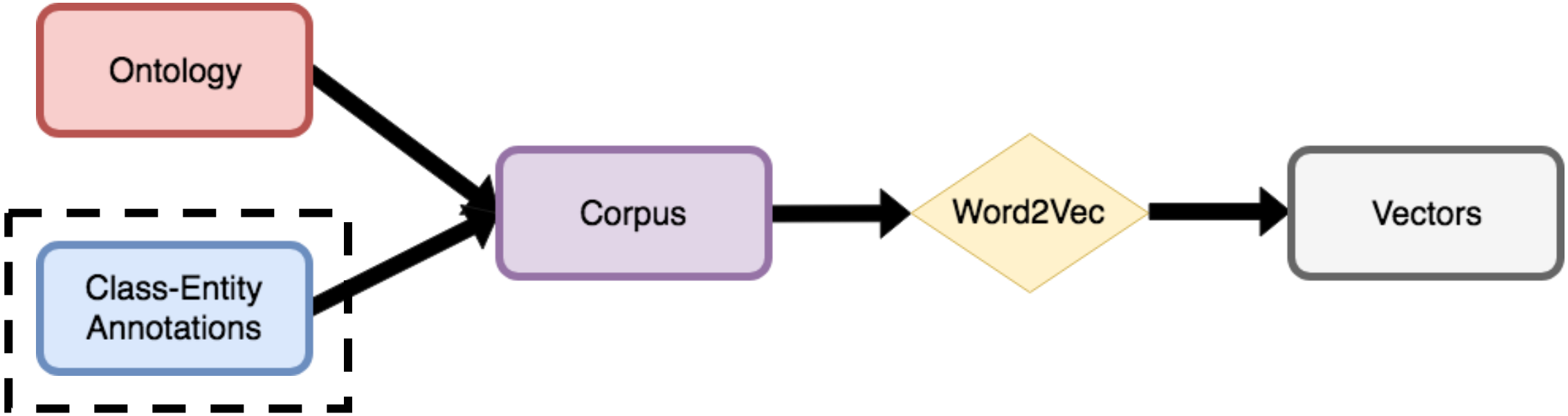
How does it work ?



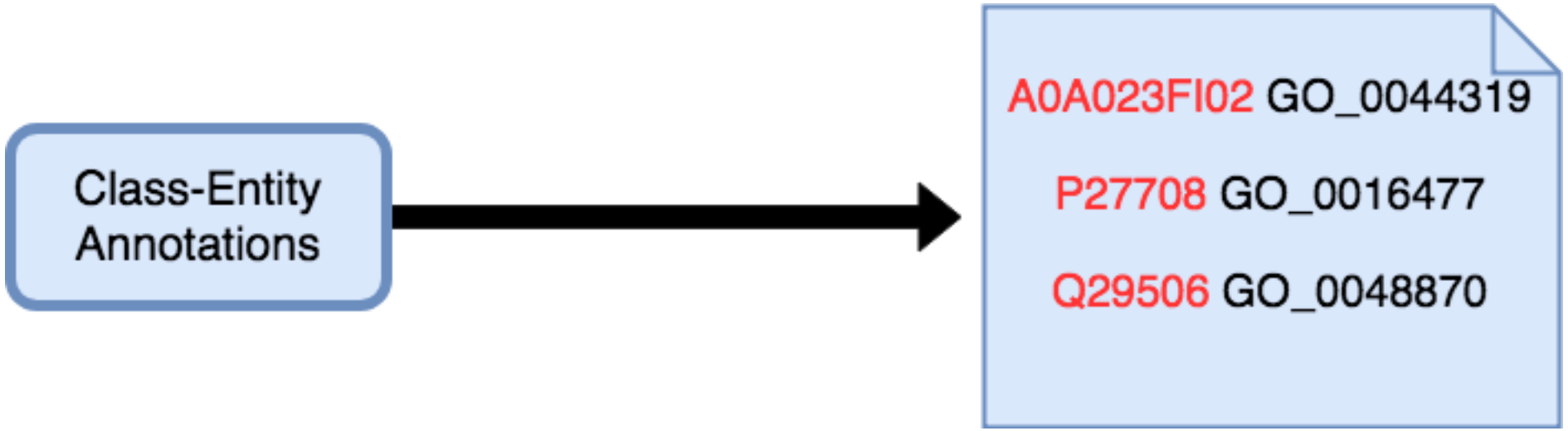
Ontology Information



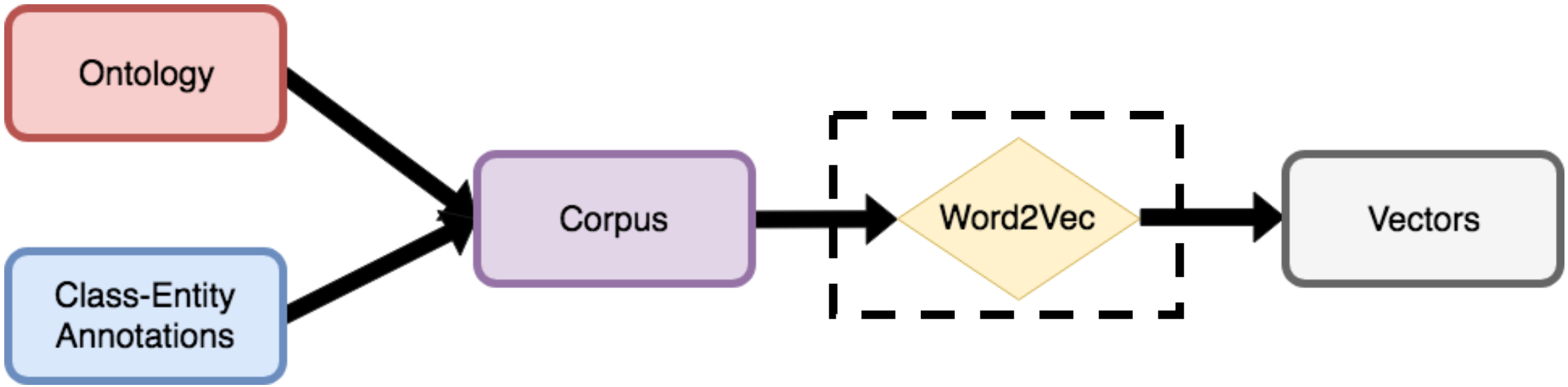
How does it work ?



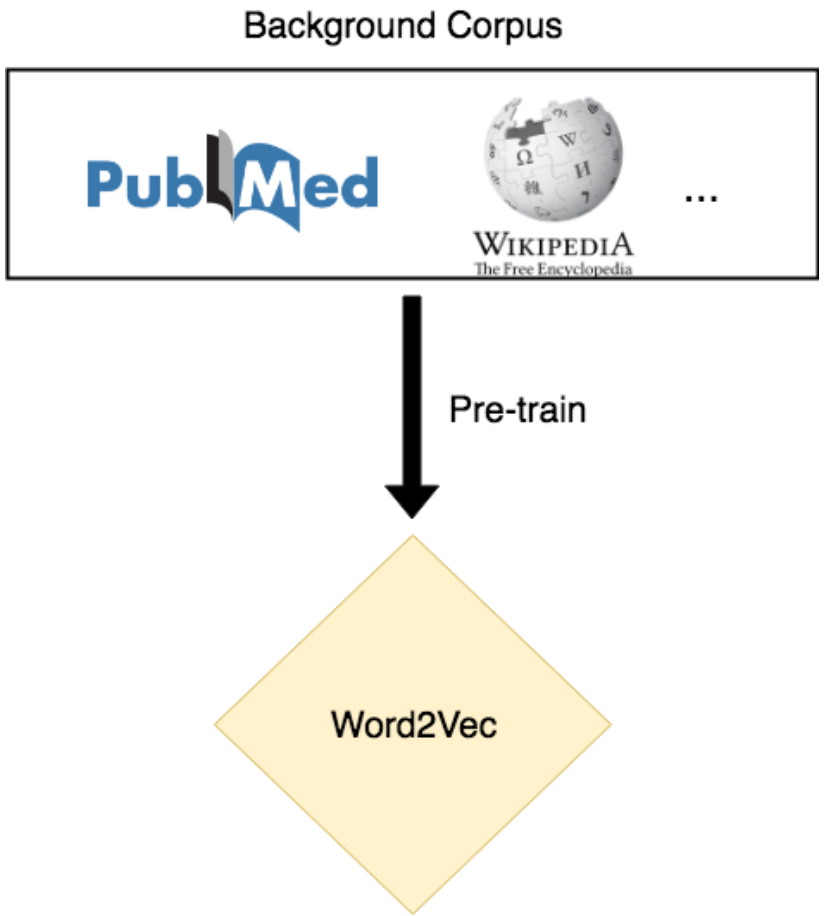
Annotation Information



How does it work ?



Word2Vec


Background Corpus
PubMed
WIKIPEDIA
The Free Encyclopedia
...
Pre-train
Word2Vec

How to use OPA2Vec

- Mandatory input:
  - Owl Ontology
  - Class-entity annotations (e.g.: GO-protein annotation)
- Optional parameters:
  - Metadata annotations: labels, description, synonyms … (default value: all)
    - e.g : <http://purl.obolibrary.org/obo/IAO_0000115>
  - Vector size (default 200)
  - Word2vec parameters:
    - Model: cbow or sg (default value: sg)
    - Window size (default value: 5)
    - Mincount (default value: 25)
    - Pre trained model: Corpus used to pre-train Word2Vec for some background information (Have to be pre-trained already).
    - List of entities for which you want to get the vector representation (default: all classes and entities)

How to run OPA2Vec ?

- Download package from:
  - https://github.com/bio-ontology-research-group/opa2vec
  - Get your ontology (e.g. *go.owl*) and annotation file (e.g. *protein-GO annotation*).

- To run with default parameters, use command:

```
python runOPA2Vec.py go.owl SampleAssociationFile.lst
```

- You can also modify the default values as shown below:

```
python runOPA2Vec.py go.owl SampleAssociationFile.lst -embedsize 50 -windsize 10 -mincount 20 -model sg -annotations all
```