

A Case Study in Bootstrapping Ontology Graphs from Textbooks

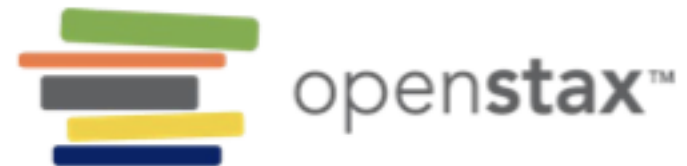
Vinay K. Chaudhri, Matt Boggess, Han L Aung

Stanford University

Debshila Basu Mallick, Andrew E. Waters, Richard G. Baraniuk

OpenStax/Rice University

Automated Knowledge Base Construction Conference, 2021



Problem

- Given an input textbook sentence, identify the classes and relationships among them

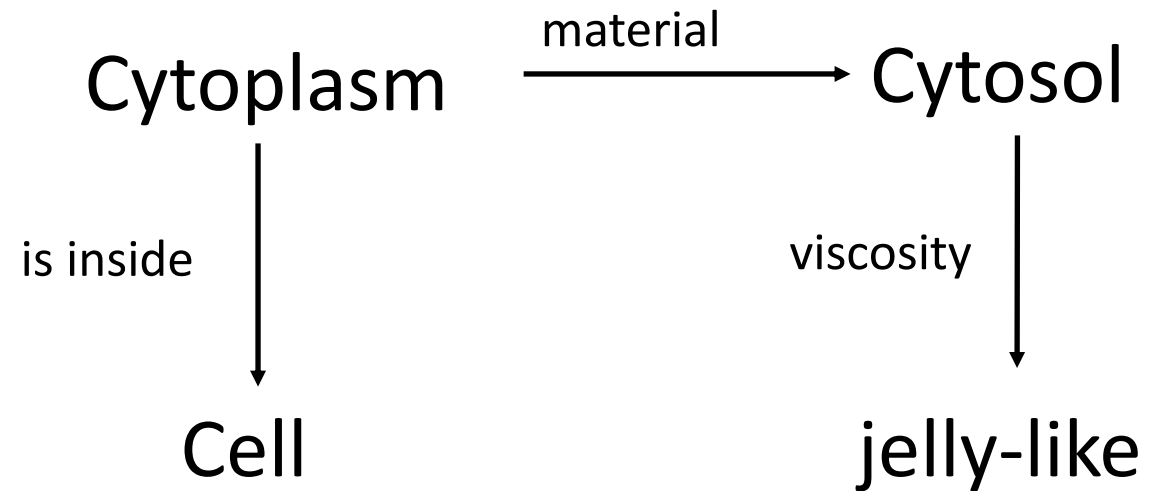
Cytoplasm consists of jelly-like cytosol within a cell



Problem

- Given an input textbook sentence, identify the classes and relationships among them

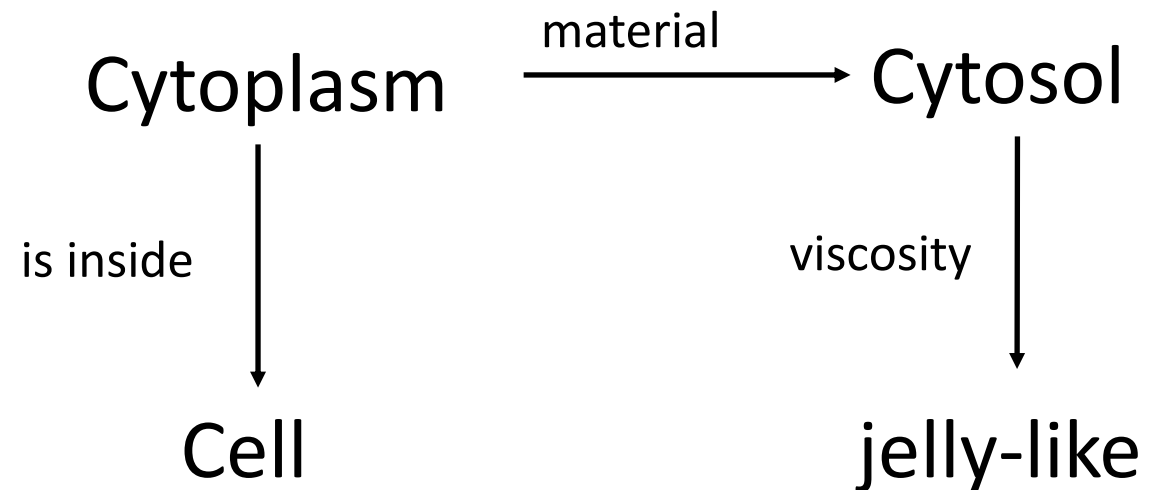
Cytoplasm consists of jelly-like cytosol within a cell



Problem

- Given an input textbook sentence, identify the classes and relationships among them

Cytoplasm consists of jelly-like cytosol within a cell



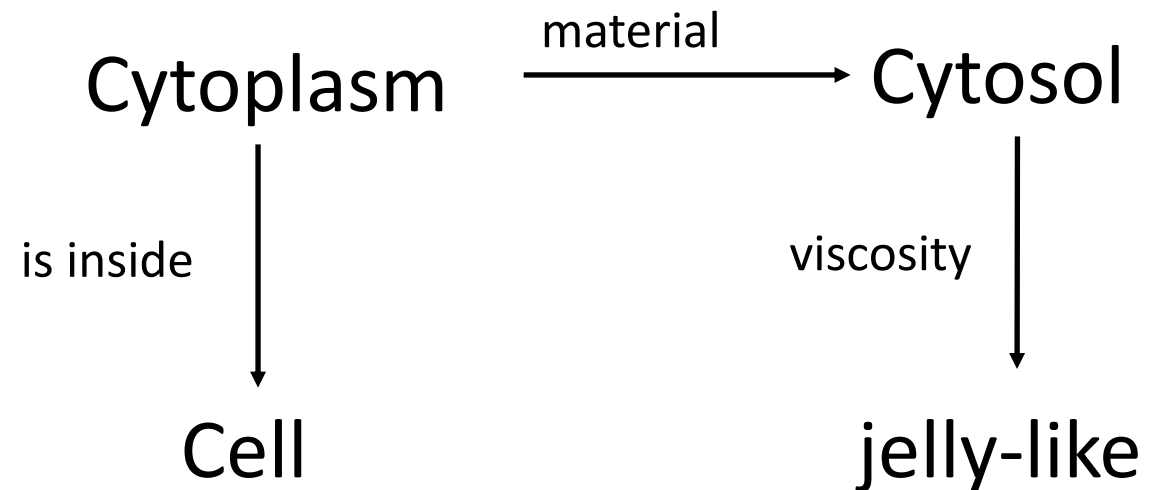
Term Extraction
Relation Extraction



Problem

- Given an input textbook sentence, identify the classes and relationships among them

Cytoplasm consists of jelly-like **cytosol** within a **cell**



Term Extraction
Relation Extraction

Where to get the training data?



Term Extraction



Input Textbook Sentences



Learned Terms



Train using
Existing
Glossary



All cells have cell membranes,
but only some have cell walls

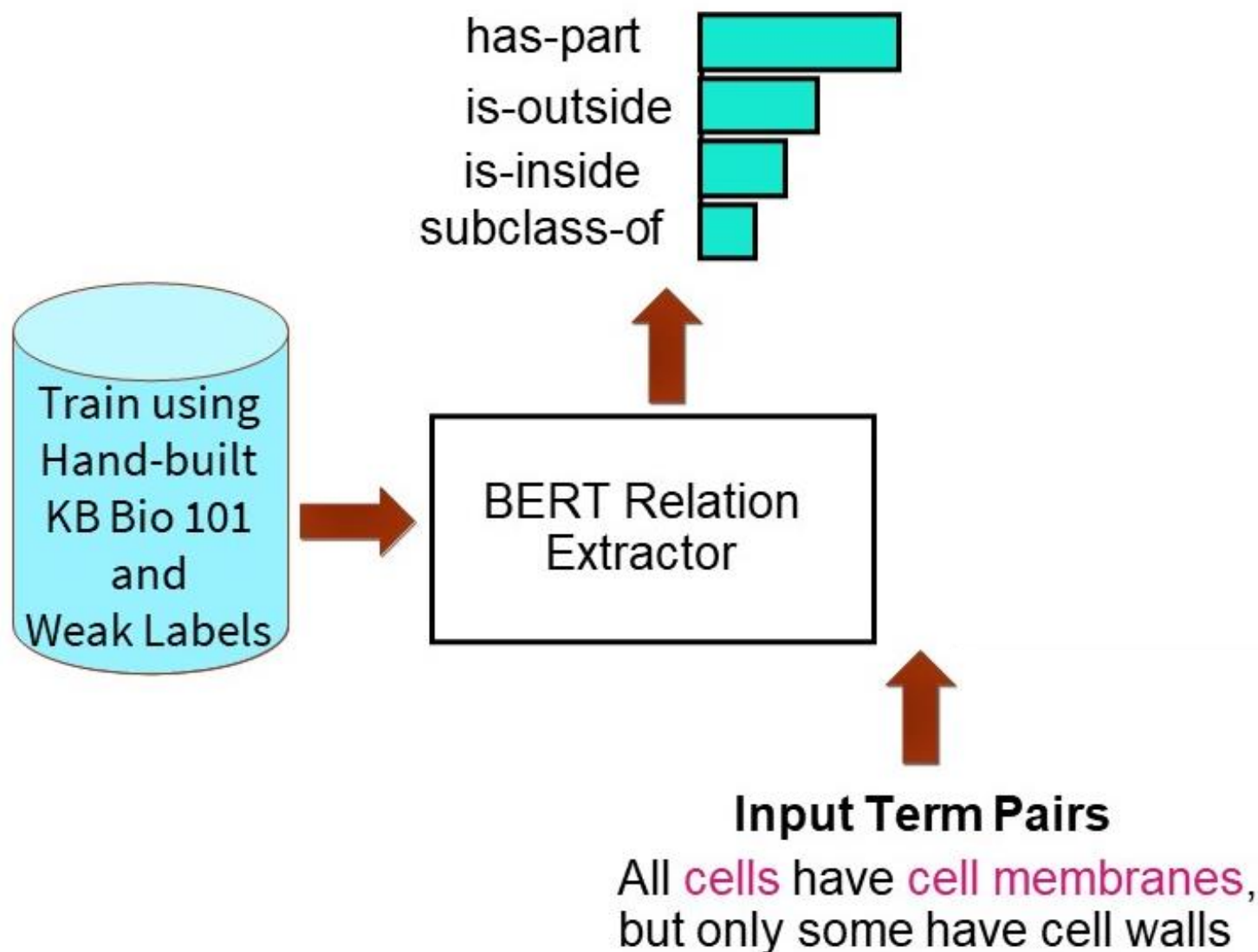


Term Extraction Results

	Term Count	AWS Comprehend F1	Term Extractor F1
Bio2e 4.2	24	0.23	0.48
Bio 2e 10.2	47	0.32	0.49
Life Chapter 39	596	0.41	0.55
Overall		0.32	0.51



Relation Extraction



Relation Extraction Results

	Precision	Recall	F1
subclass	0.32	0.88	0.47
superclass	0.53	0.76	0.62
synonym	0.25	1.00	0.42
has part / region	0.15	0.50	0.22
has part of / region of	0.17	0.80	0.28



Relation Extraction

The nucleus resides inside the cytoplasm

[Click here for the full content \(will open a new tab\)](#)

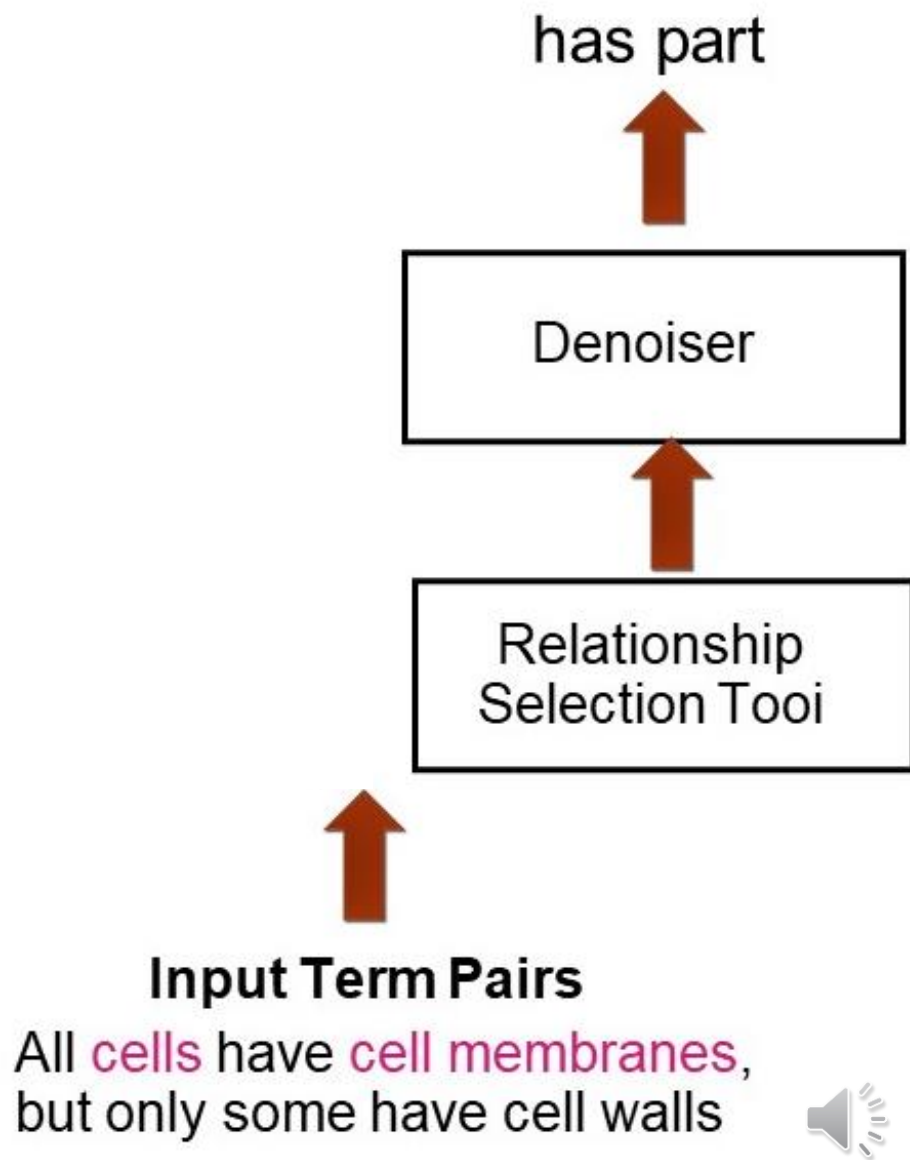
Good job! Now we need to select the correct subclass relationship from the new list. Here we would choose 'is inside of' on the right.

What type of relationship exists between cytoplasm nucleus?

nucleus

cytoplasm

- Flip Term Order
- is located at
- is inside of
- resides against
- I am not sure
- Define a new relationship
- Redo Family Selection



Relation Selection Results

Text Book Section	Precision when either SME agreed
Biology 4.2	0.98
Biology 10.2	0.96
Psychology 14.1	0.98



Train using
Existing
Glossary



Term Extraction



Input Textbook Sentences

BERT Term
Extractor

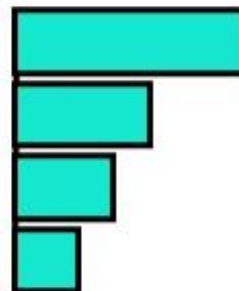
Learned Terms

All **cells** have **cell membranes**,
but only some have **cell walls**

Train using
Hand-built
KB Bio 101
and
Weak Labels

BERT Relation
Extractor

has-part
is-outside
is-inside
subclass-of



Relation Extraction

has part

Denoiser

Relationship
Selection Tool

Input Term Pairs

All **cells** have **cell membranes**,
but only some have **cell walls**



Conclusions and Future Work

- Automation is useful in bootstrapping term extraction and taxonomy building
- Requires more training to do well on relation selection

Code and Data available at:

<https://openstax.github.io/research-kg-learning/akbc-2021/>

