Information Extraction
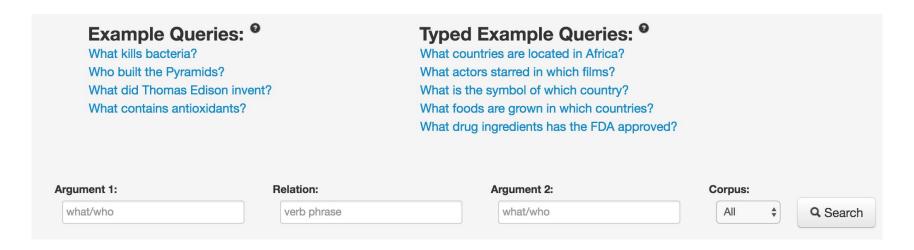
ReVerb/TextRunner by UW

Demo: http://openie.allenai.org/

Software: http://reverb.cs.washington.edu/

Example Queries: ❓
What kills bacteria?
Who built the Pyramids?
What did Thomas Edison invent?
What contains antioxidants?

Typed Example Queries: ❓
What countries are located in Africa?
What actors starred in which films?
What is the symbol of which country?
What foods are grown in which countries?
What drug ingredients has the FDA approved?

Argument 1:
what/who

Relation:
verb phrase

Argument 2:
what/who

Corpus:
All ⇕

🔍 Search

ReVerb/TextRunner by UW

Demo: http://openie.allenai.org/

Software: http://reverb.cs.washington.edu/

1. Open domain
2. Extract all relation tuples (not restricted to certain relations)
3. More about "string surface"

MITIE

Official Page: https://github.com/mit-nlp/MITIE

1. Can be run as command line or as library
2. Support multiple programming languages

Reference:

[1] Davis E. King. Dlib-ml: A Machine Learning Toolkit. Journal of Machine Learning Research 10, pp. 1755-1758, 2009.

[2] Paramveer Dhillon, Dean Foster and Lyle Ungar, Eigenwords: Spectral Word Embeddings, Journal of Machine Learning Research (JMLR), 16, 2015.

[3] T. Joachims, T. Finley, Chun-Nam Yu, Cutting-Plane Training of Structural SVMs, Machine Learning, 77(1):27-59, 2009.

MITIE

Official Page: https://github.com/mit-nlp/MITIE

Input: A Pegasus Airlines plane landed at an Istanbul airport Friday after a passenger "said that there was a bomb on board" and wanted the plane to land in Sochi, Russia, the site of the Winter Olympics, said officials with Turkey's Transportation Ministry.

Middle: A [ORGANIZATION Pegasus Airlines] plane landed at an [LOCATION Istanbul] airport Friday after a passenger " said that there was a bomb on board " and wanted the plane to land in [LOCATION Sochi] , [LOCATION Russia] , the site of the Winter Olympics , said officials with [LOCATION Turkey] 's [ORGANIZATION Transportation Ministry] .

Final: Number of named entities detected: 23

```
Now look for binary relations.
relation type: location.location.contains
Russia   #   Sochi
```

Stanford openIE

Open information extraction (open IE) refers to the extraction of relation tuples, typically binary relations, from plain text. The central difference is that the schema for these relations does not need to be specified in advance; typically the relation name is just the text linking two arguments.

For example, Barack Obama was born in Hawaii would create a triple (Barack Obama; was born in; Hawaii), corresponding to the open domain relation was-born-in(Barack-Obama, Hawaii).

Usage: `java -mx1g -cp "*" edu.stanford.nlp.naturalli.OpenIE  /path/to/file1  /path/to/file2`