# Annotation via DBpedia

## DBpedia: What is it?

From the DBpedia homepage we read the following (https://wiki.dbpedia.org/about):

> DBpedia is a crowd-sourced community effort to extract structured content from the information created in various Wikimedia projects. This structured information resembles an open knowledge graph (OKG) which is available for everyone on the Web. A knowledge graph is a special kind of database which stores knowledge in a machine-readable form and provides a means for information to be collected, organised, shared, searched and utilised. Google uses a similar approach to create those knowledge cards during search. We hope that this work will make it easier for the huge amount of information in Wikimedia projects to be used in some new interesting ways.

## OWL: What is it?



DBpedia ⊘ Browse using ▾ ▤ Formats ▾     ⎘ Faceted Browser   ⎘ Sparql Endpoint
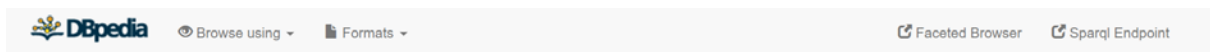
## About: Web Ontology Language

An Entity of Type : Communication100033020, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

The Web Ontology Language (OWL) is a family of knowledge representation languages for authoring ontologies. Ontologies are a formal way to describe taxonomies and classification networks, essentially defining the structure of knowledge for various domains: the nouns representing classes of objects and the verbs representing relations between the objects. Ontologies resemble class hierarchies in object-oriented programming but there are several critical differences. Class hierarchies are meant to represent structures used in source code that evolve fairly slowly (typically monthly revisions) whereas ontologies are meant to represent information on the Internet and are expected to be evolving almost constantly. Similarly, ontologies are typically far more flexible as they are meant to repres

## The DBpedia annotator GUI

The NLP Suite provides a convenient GUI to annotate a text via DBpedia entries.
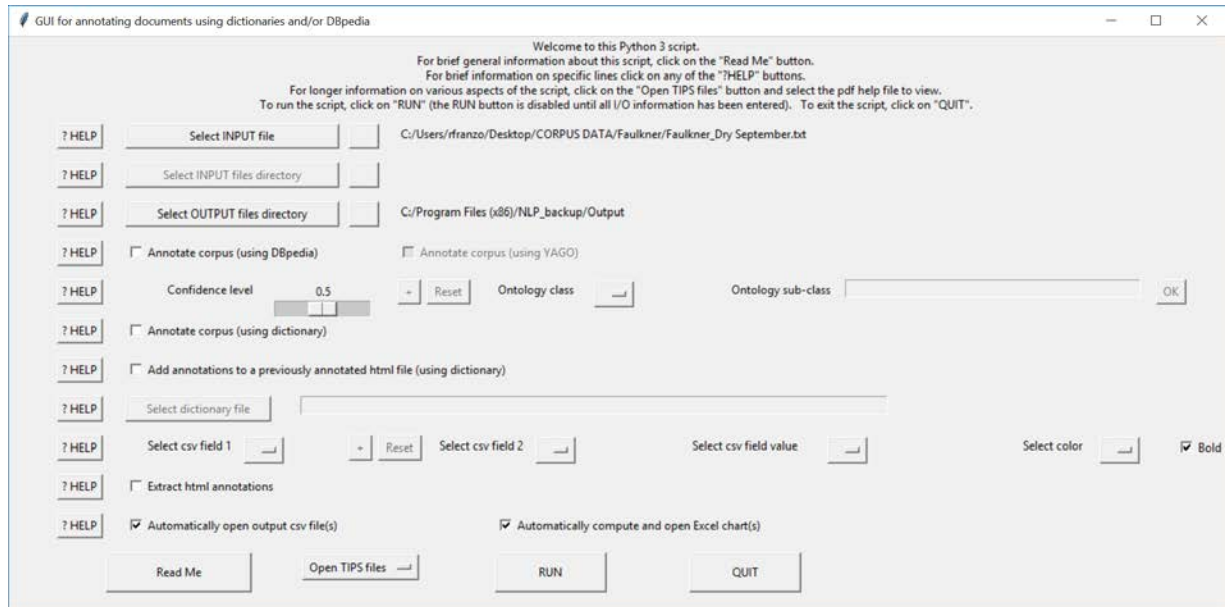


## DBpedia demo

DBpedia offers an online demo at https://www.dbpedia-spotlight.org/demo/. You can use the website to test any texts that you may wish. Just replace the demo text ("First documented in the 13th century, Berlin was the capital of the Kingdom of Prussia…") with our text, select the DBpedia ontology class you wish to use then click Annotate to see the results.

## Confidence level & n-best candidates: What are they?

The DBpedia Spotlight demo offers 2 parameters: **confidence** and **n-best candidates**. What are they?

### *Confidence*

The slider widget allows you to choose a Confidence value between 0 and 1. Assigning the value 0 to this parameter means that DBpedia detects every possible mention of any word in the text, while the value 1 keeps only mentions for which DBpedia is sure of the correctness of the linked entity. To determine which mentions are retained, DBpedia Spotlight computes the difference between the score of the best candidate entity e1 and the second-best candidate entity e2 for the same mention m. If their scores are close, DBpedia Spotlight returns the annotation m → e1 only if the value assigned to the confidence parameter is low. Put simply, the bigger the confidence value, the greater the score difference must be in order to keep the best candidate. In practice, **the higher the Confidence value, the lower the probability that DBpedia will actually annotate your text. The recommended value is 0.5.**

### *n-best candidates*

If an item in your text can be found in more than one ontology class (e.g., Berlin being both in Germany and other locations). When DBpedia annotates Berlin with the n-best candidates ticked, Berlin will be annotated with a down arrow. When you click on the arrow DBpedia will display all the ontology classes that it found for Berlin (or any other annotated expression).

### *Spotlight is VERY spotty…*

**Don't be surprised if you run the DBpedia Spotlight demo and you get nothing in return even when using their demo text and selecting all annotations with confidence value 0**
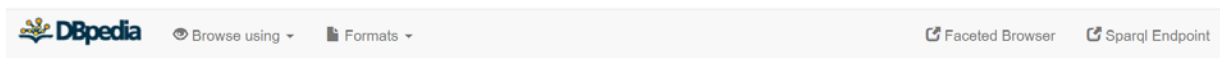
**and n-best candidates ticked. That is unfortunately a common problem with DBpedia regardless of the browser you use.**

<span style="color:red">**The output html annotated file**</span>

William Faulkner's short story *Dry September* annotated using the default classes will look like this:



Clicking on any of the annotated entries, for example on WILLIAM FAULKNER, will take us to the DBpedia website



<span style="color:red">**Why isn't 'moon' annotated?**</span>

In the last sentence of Faulkner's *Dry September* "The dark world seemed to lie stricken beneath the cold moon and the lidless stars." 'moon' is not annotated, regardless of confidence level. That is the case even when I just enter the class 'Planet'. Why?

## About: planet

An Entity of Type : Class, from Named Graph : http://dbpedia.org/resource/classes#, within Data Space : dbpedia.org

| Property | Value |
| --- | --- |
| rdf:type | ▪ owl:Class |
| rdfs:isDefinedBy | ▪ http://dbpedia.org/ontology/ |
| rdfs:label | ▪ planet (en) |
| rdfs:subClassOf | ▪ dbo:CelestialBody |
| owl:equivalentClass | ▪ wikidata:Q634 |
| wdrs:describedby | ▪ dbo:data/definitions.ttl |

## About: celestial body

An Entity of Type : Class, from Named Graph : http://dbpedia.org/resource/classes#, within Data Space : dbpedia.org

| Property | Value |
| --- | --- |
| rdf:type | ▪ owl:Class |
| rdfs:isDefinedBy | ▪ http://dbpedia.org/ontology/ |
| rdfs:label | ▪ celestial body (en) <br> ▪ corpo celeste (it) |
| rdfs:subClassOf | ▪ dbo:Place |
| wdrs:describedby | ▪ dbo:data/definitions.ttl |

## About: Moon

An Entity of Type : planet, from Named Graph : http://dbpedia.org, within Data Space : dbpedia.org

The Moon is Earth's only permanent natural satellite. It is the fifth-largest natural satellite in the Solar System, and the largest among planetary satellites relative to the size of the planet that it orbits (its primary). It is the second-densest satellite among those whose densities are known (after Jupiter's satellite Io). The average distance of the Moon from the Earth is 384,400 km (238,900 mi), or 1.28 light-seconds.

**Notice how 'Moon' is capitalized in the DBpedia database under the Entity of Type planet. If you capitalize moon it will be annotated.**

*Forbidden/Permission denied: Why do I get this?*

If your output file only contains the following message, don't worry. The DBpedia server was having a bad moment. This happens to the best of us, including DBpedia. Don't get discouraed. Simply repeat your query.

# Forbidden

You don't have permission to access this resource.

---

*Apache/2.4.25 (Debian) Server at model.dbpedia-spotlight.org Port 80*

## *Output filenames*

You can recognize a file that has been annotated via DBpedia by a specific tag:

NLP_DBpedia_annotated_Faulkner_Dry September.html

## Further readings: More TIPS files

TIPS_NLP_Annotator.pdf
TIPS_NLP_Annotator dictionary.pdf
TIPS_NLP_Annotator extractor.pdf