

Annotation via DBpedia

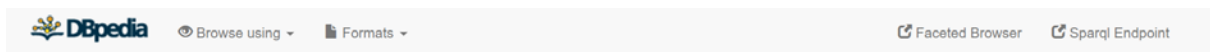
DBpedia: What is it?	1
OWL: What is it?	1
The DBpedia annotator GUI	2
DBpedia demo.....	2
Confidence level & n-best candidates: What are they?.....	3
Confidence.....	3
n-best candidates.....	3
Spotlight is VERY spotty... ..	3
The output html annotated file	4
Why isn't 'moon' annotated?	4
Forbidden/Permission denied: Why do I get this?	5
Output filenames	6
Further readings: More TIPS files.....	6

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?



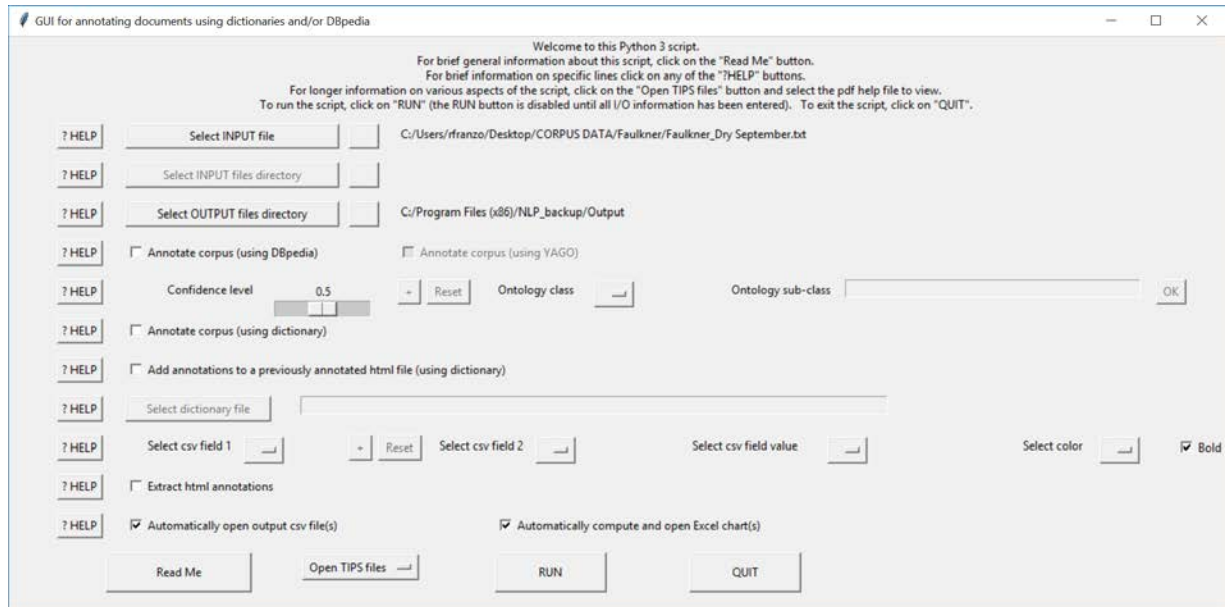
About: Web Ontology Language

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: Language:

☐ n-best candidates

First documented in the 13th century, Berlin was the capital of the Kingdom of Prussia (1701–1918), the German Empire (1871–1918), the Weimar Republic (1919–33) and the Third Reich (1933–45). Berlin in the 1920s was the third largest municipality in the world. After World War II, the city became divided into East Berlin -- the capital of East Germany -- and West Berlin, a West German exclave surrounded by the Berlin Wall from 1961–89. Following German reunification in 1990, the city regained its status as the capital of Germany, hosting 147 foreign embassies.

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 e_1 and the second-best candidate entity e_2 for the same mention m . If their scores are close, DBpedia Spotlight returns the annotation $m \rightarrow e_1$ 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

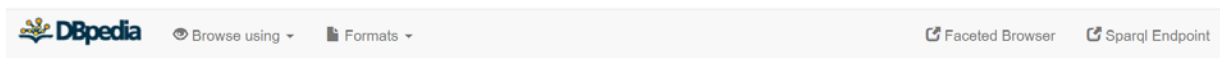
and n-best candidates ticked. That is unfortunately a common problem with DBpedia regardless of the browser you use.

The output html annotated file

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

[WILLIAM FAULKNER](#) Dry September THROUGH THE BLOODY September twilight, aftermath of sixty-two rainless days, it had gone like a fire in dry [grass](#): the rumor, the story, whatever it was. Something about Miss Minnie Cooper and a Negro. Attacked, insulted, frightened: none of them, gathered in the barber shop on that Saturday evening where the [ceiling fan](#) stirred, without freshening it, the vitiated air, sending back upon them, in recurrent surges of stale [pomade](#) and lotion, their own stale breath and odors, knew exactly what had happened. Except it wasn't Will Mayes, a barber said. He was a man of middle age; a thin, sand-colored man with a mild face, who was [shaving](#) a client. I know Will Mayes. He's a good [nigger](#). And I know Miss Minnie Cooper, too. What do you know about her? a second barber said. Who is she? the client said. A young girl? No, the barber said. She's about forty, I reckon. She ain't married. That's why I don't believe. Believe, hell! a hulking youth in a sweat-stained [silk](#) shirt said. Wont you take a white woman's word before a [nigger's](#)? I don't believe Will Mayes did it, the barber said. I know Will Mayes. Maybe you know who did it, then. Maybe you already got him out of town, you damn nigger-lover. I don't believe anybody did anything. I don't believe anything happened. I leave it to you fellows if them ladies that get old without getting married don't have notions that a man can't Then you are a hell of a white man, the client said. He moved under the cloth. The youth had sprung to his feet. You don't? he said. Do you accuse a white woman of lying? The barber held the [razor](#) poised above the half-risen client. He did not look around. It's this durn weather, another said. It's enough to make a man do anything. Even to her. Nobody laughed. The barber said in his mild, stubborn tone: I ain't accusing nobody of nothing. I just know and you fellows know how a woman that never You damn nigger-lover! the youth said. Shut up, [Butch](#), another said. We'll get the facts in plenty of time to act. Who is? Who's getting them? the youth said. Facts, hell! You're a fine white man, the client said. Ain't you? In his frothy beard he looked like a [desert rat](#) in the moving pictures. You tell them, [Jack](#), he said to the youth. If there ain't any white men in this town, you can count on me, even if I ain't only a drummer and a stranger. That's right, boys, the barber said. Find out the truth first. I know Will Mayes. Well, by [God](#)! the youth shouted. To think that a white man in this town. Shut up, [Butch](#), the second speaker said. We got plenty of time. The client sat up. He looked at the speaker. Do you claim that anything excuses a [nigger](#) attacking a white woman? Do you mean to tell me you are a white man and you'll stand for it? You better go back North where you came from. The South don't want your kind here. North what? the second said. I was born and raised in this town. Well, by [God](#)! the youth said. He looked about with a strained, baffled gaze, as if he was trying to remember what it was he wanted to say or to do. He drew his sleeve across his [sweating](#) face. Damn if I'm going to let a white woman. You tell them, [Jack](#), the drummer said. By [God](#), if they. The screen door crashed open. A man stood in the floor, his feet apart and his heavy-set body poised easily. His white shirt was open at the throat; he wore a felt hat. His hot, bold glance swept the group. His name was McLendon. He had commanded troops at the front in [France](#) and had been decorated for valor. Well, he said, are you going to sit there and let a black son [rape](#) a white woman on the streets of [Jefferson](#)? [Butch](#) sprang up again. The [silk](#) of

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



About: William Faulkner

An Entity of Type : [person](#), from Named Graph : <http://dbpedia.org>, within Data Space : [dbpedia.org](#)

William Cuthbert Faulkner (/ˈfɔːlkənər/, September 25, 1897 – July 6, 1962) was an American writer and Nobel Prize laureate from Oxford, Mississippi. Faulkner wrote novels, short stories, a play, poetry, essays, and screenplays. He is primarily known for his novels and short stories set in the fictional Yoknapatawpha County, based on Lafayette County, Mississippi, where he spent most of his life.

Why isn't 'moon' annotated?

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	<ul style="list-style-type: none"> owl:Class
rdfs:isDefinedBy	<ul style="list-style-type: none"> http://dbpedia.org/ontology/
rdfs:label	<ul style="list-style-type: none"> planet (en)
rdfs:subClassOf	<ul style="list-style-type: none"> dbo:CelestialBody
owl:equivalentClass	<ul style="list-style-type: none"> wikidata:Q634
wdrs:describedby	<ul style="list-style-type: none"> 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	<ul style="list-style-type: none"> owl:Class
rdfs:isDefinedBy	<ul style="list-style-type: none"> http://dbpedia.org/ontology/
rdfs:label	<ul style="list-style-type: none"> celestial body (en) corpo celeste (it)
rdfs:subClassOf	<ul style="list-style-type: none"> dbo:Place
wdrs:describedby	<ul style="list-style-type: none"> 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