

Text Annotation

Annotation: What is it?	1
Why annotating?	1
The Annotator GUI	1
HTML: What is it?	2
HTML file: What does it look like in a web browser?.....	2
HTML file: What does it look like in a text editor?.....	2
DBpedia and dictionary annotation: What is the difference?	3
Output filenames	3
Further readings: More TIPS files.....	4

Annotation: What is it?

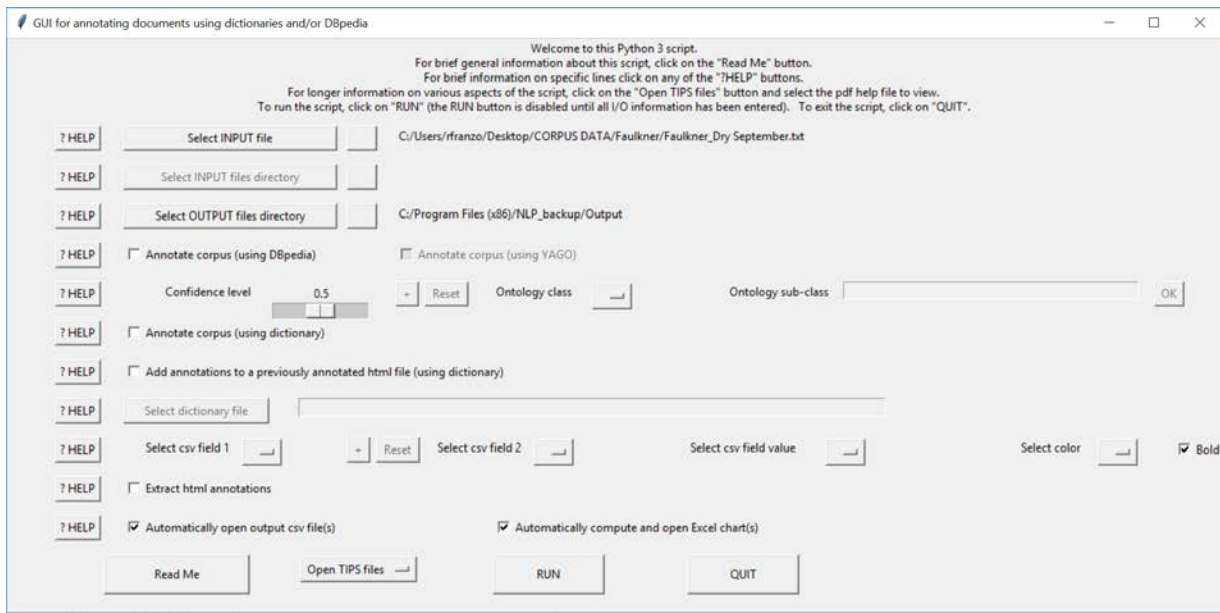
Annotating a text typically involves highlighting or underlining key pieces of text and making notes in the margins of the text. In NLP, annotation is carried out automatically, rather than by hand with a highlighter or a pencil. Notes are found on the web, rather than in the margins, according to a knowledge base (e.g., DBpedia or YAGO).

Why annotating?

You typically manually annotate a text (sometimes in different colors) to stress the most important points mentioned in the text, to facilitate a quick re-reading and recalling of the text. Furthermore, color highlighting may help to suggest relationships. NLP annotation has the same goals as manual annotation, except that it is all done automatically by a computer algorithm. If the former is perhaps an example of *close reading*, the latter is an example of *distant reading*.

The Annotator GUI

The NLP Sute provides a number of Python 3 scripts for annotating a text in HTML format, either with DBpedia entries or with dictionaries and for extracting annotations from an HTML file.



HTML: What is it?

Both dictionary and DBpedia annotators create HTML files in output. Hypertext Markup Language (HTML) is the standard markup language for documents designed to be displayed in a web browser.

HTML file: What does it look like in a web browser?

Depending upon what choose to annotate (Things, Unknown) and the color you choose for the annotations (blue), this is what the beginning of Faulkner's short story *Dry September* looks like (annotated with DBpedia option).

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

HTML file: What does it look like in a text editor?

You can open an html file in a text editor to reveal the code behind it. Here is what it looks like:

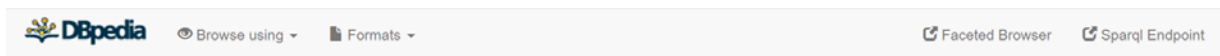
```

<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.01 Transitional//EN" "http://www.w3.org/TR/html4/loose.dtd">
<html>
<head>
<title>DBpedia Spotlight annotation</title>
<meta http-equiv="Content-type" content="text/html; charset=UTF-8">
</head>
<body>
<div>
<span style="color: blue; font-weight: bold">WILLIAM FAULKNER</span>
Dry September THROUGH THE BLOODY September twilight, aftermath of sixty-two rainless days, it had gone like a fire in dry <span style="color: blue; font-weight: bold">grass</span> the rumor, the story, whatever it was. Something about Miss Minnie Cooper and a Negro. Attacked, insulted, frightened: none of them, gatheblue in the barber shop on that Saturday evening where the <span style="color: blue; font-weight: bold">ceiling fan</span> stirblue, without freshening it, the vitiated air, sending back upon them, in recurrent surges of stale <span style="color: blue; font-weight: bold">pomade</span> 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-colobblue man with a mild face, who was shaving a client. I know Will Mayes. He's a good <span style="color: blue; font-weight: bold">nigger</span>. 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 <span style="color: blue; font-weight: bold">silk</span> shirt said. Wont you take a white woman's word before a <span style="color: blue; font-weight: bold">nigger</span>'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.
</div>
</body>
</html>

```

DBpedia and dictionary annotation: What is the difference?

Just looking at a DBpedia or dictionary annotated html file, there will be little difference: a bunch of words in the text will be highlighted. But the words annotated with DBpedia will be underlined, they are hiperlinks to a web entry in the DBpedia knowledge base. Thus, if I click on WILLIAM FAULKNER I get the following entry:



William Cuthbert Faulkner (/ˈfoːlknər/, 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.

The NLP note on the margin...

The words annotated via a dictionary (e.g., annotating all NER PERSON in the text) would **not** be underlined. They would be highleted but not underlined; there is no hiperlink to a knowledge base. Nothing will happen when you click on the highlighted word.

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 Negro. Attacked, insulted, frightened: none of them, gatheblue in the barber shop on that Saturday evening where the **ceiling fan** stirblue, 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-colobblue 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.

Output filenames

You can recognize a file that has been annotated by specific tags:

NLP_DBpedia_annotated_Faulkner_Dry September.html
 NLP_dict_annotated_Faulkner_Dry September.html

In the first case, the filename tells me that Faulkner's *Dry September* was annotated via DBpedia.

In the second case, the filename tells me that Faulkner's *Dry September* was annotated dictionary.

html files that have undergone multiple dictionary annotations are similarly specifically marked.

NLP_DBpedia_annotated_dict_annotated_Faulkner_Dry September.html
NLP_DBpedia_annotated_multiDict_annotated_Faulkner_Dry September.html
NLP_multiDict_annotated_Faulkner_Dry September.html

In the first case, the filename tells me that the html DBpedia annotated Faulkner's *Dry September* has been re-annotated via a dictionary.

In the second case, that same html file has been re-annotated multiple times via a dictionary.

In the third case, Faulkner's *Dry September* originally annotated via a dictionary has then been re-annotated multiple times via a dictionary.

Further readings: More TIPS files

TIPS_NLP_Annotator DBpedia.pdf
TIPS_NLP_Annotator dictionary.pdf
TIPS_NLP_Annotator extractor.pdf