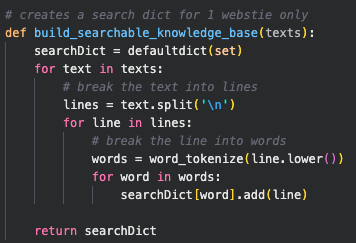
Dmitrii Obideiko

Suraj Janakiraman

**How we created a knowledge base:**

We created a knowledge base by first clearing up all the text in files. For example all empty lines were removed as well as any spaced between the lines. Then all lines from every text were combined into one big list of lines. A dictionary of sets was used to keep all occurrences in a line for a specific word. (The key was a word and the value was a set of lines where that word appeared.)



Sample Screenshots of knowledge base (text):



You will see at the bottom that the words “form” and “read are part of the 10 terms listed.



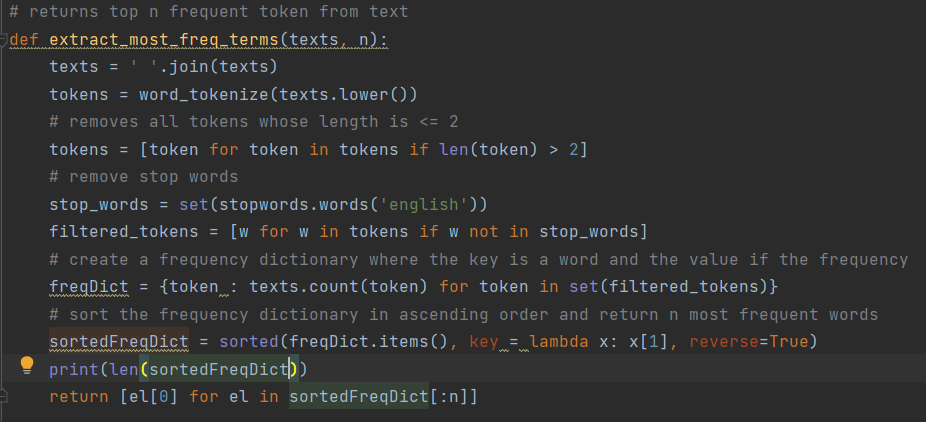
You will see at the bottom that the word “ work” is one of the 10 terms listed. There are multiple usages of the word “work” in the knowledge base.



Notice the word sign most typically appears in the word **“design.”** You will see at the bottom that the word “sign” is one of the 10 terms listed.

**Indication of the top 10 terms**

We created a function that gets the mostly frequent words/tokens from each text in the list of urls.



**Output:**

['pro', 'ted', 'per', 'min', 'art', 'act', 'low', 'one', 'form', 'men', 'eat', 'use', 'tim', 'work', 'sign', 'app', 'age', 'end', 'able', 'ming', 'format', 'king', 'formation', 'read', 'inc']

**From this function, we can choose the top 10 terms which can be classified as actual words:**

1. form
2. act
3. work
4. sign
5. app
6. age
7. end
8. able
9. format
10. read

**Now we create a dialog based on the knowledge base:**

* our story needs the best idea
* think about them
* share ideas
* easy our story is the best