

Methodology

1. We'll take the input from the user.
2. Preprocess the Operand String.
3. Preprocess the Operator String.
4. Start a loop for the number of operators and process the preprocessed query words left to right.
5. Print the number of documents matched and Number of comparisons used.

Preprocessing

- Load stories dataset having 467 files
- Load the files location using `os.path.join`
- Iterate through each file
 - Removing the punctuations, symbols, numbers.
 - Tokenizing the words.
 - Iterating through the tokens
 - Removing the stopwords
 - Correcting the elongated words
 - Removing repeated words

Creating InvertedIndex

While going through each file and preprocessing it. We'll take the words and if the word is not already present as a key in `invertedIndex` create a new posting list with current document ID otherwise add the current document ID at the appropriate position in the already sorted posting list of the word

Then insert the length of the posting list for every word as it's first posting list entry.

Assumptions

All the 467 files are present in the folder Stories.

The operators are:

1. AND
2. OR
3. NOT
4. AND NOT
5. OR NOT