

IR Assignment #1 (Due 4th Sep)

Construct a positional index structure in Java (or python). You have to read multiple (at least 5) files from a folder and construct inverted index using the words as terms. You can use Java Collection Classes. The format of each posting is of the form:

Term: { docid, {pos1, pos2.....} },
 { docid, {pos1, pos2.....} }

See examples in pages 38 and 41.

Implement the following search queries on your index structure. (document IDs are just for representation in the following examples, they can differ in your answer)

Note: All the words are stored in small caps. Convert all the queries to small alphabets. Remove all special characters.

(a). old // doc ids of docs containing the term old

Query, "old" occurs at Document ID(s) -

- 1: Frogs that rode a Snake.txt
- 2: The Greedy Cobra and the King of Frogs.txt
- 3: The Old Man, his Young Wife and The Thief.txt
- 4: The Snake and The Foolish Frogs.txt
- 6: The War of Crows and Owls.txt

(b) snake AND frog // doc ids of docs containing both the terms

Query, "snake AND frog" occurs at Document ID(s) -

- 1: Frogs that rode a Snake.txt
- 4: The Snake and The Foolish Frogs.txt

(c) snake OR frog

Query, "snake OR frog" occurs at Document ID(s) -

- 1: Frogs that rode a Snake.txt
- 2: The Greedy Cobra and the King of Frogs.txt
- 4: The Snake and The Foolish Frogs.txt
- 5: The Tale of Two Snakes.txt

(d) "There was a" // phrase search

Query, "There was a" occurs at Document ID(s) -

- 3: The Old Man, his Young Wife and The Thief.txt
- 5: The Tale of Two Snakes.txt
- 6: The War of Crows and Owls.txt
- 7: The Wedding of the Mouse.txt

(e) **"There was a" AND "One day"** // phrase search

Query, "There was a AND One day" occurs at Document ID(s) -

5: The Tale of Two Snakes.txt

7: The Wedding of the Mouse.txt

(f) **Why /2 you** // why occurring within 2 words of you

Query, "Why /2 you" occurs at Document ID(s) -

1: Frogs that rode a Snake.txt

2: The Greedy Cobra and the King of Frogs.txt

5: The Tale of Two Snakes.txt

Submission : In a word document, summarize your results and submit along with source file in the shared drive in the following location:

1. Go to **"Submissions"** folder on the shared drive.
2. Create a folder for your team. Eg. **Team1**
3. Under *Team1*, create **Assignment1** folder.
4. Copy your results and the source files in that folder