

Information Retrieval

Assignment # 1

Assignment # 1 is due March 22 12:00 noon

Groups of 3

All three must be able to describe the WHOLE program

Ramadan is great for programming.....

Write a java code that

- 1- Read 10 text files.
- 2- Build the inverted index for those 10 file
- 3- Read one word and list all files containing the word.

The following is not optional:

- a. The posting list should be implemented as a **linked list of java objects** called **“Posting”**

```
public class Posting {
    public Posting next = null;
    int docId;
    int dtf = 1; // document term frequency
    ....
```

- b. The terms can be stored in a hashmap

```
HashMap<String, Entry> index;
```

- c. Each entry contains

```
public class DictEntry {
    int doc_freq = 0; // number of documents that contain the term
    int term_freq = 0; //number of times the term is mentioned in the collection
    Posting pList = null;
```

- d. after you run and test your project perform clean and compress it
you will be given an email address to send it to.

- e. **Start working inordinately there will be no late submission.**

term	doc. freq.	→	postings lists
ambitious	1	→	2
be	1	→	2
brutus	2	→	1 → 2
capitol	1	→	1
caesar	2	→	1 → 2
did	1	→	1
enact	1	→	1
hath	1	→	2
i	1	→	1
i'	1	→	1
it	1	→	2
julius	1	→	1
killed	1	→	1
let	1	→	2
me	1	→	1
noble	1	→	2
so	1	→	2
the	2	→	1 → 2
told	1	→	2
you	1	→	2
was	2	→	1 → 2
with	1	→	2

بالتوفيق

Good luck