IR Assignment 1 Analysis File

Question 1:

- The queries which contain only the operation OR NOT will have the highest number of comparisons as the doc-id list obtained after applying NOT will be very long.
- Also, the queries which have other operators along with OR NOT will also be having high number of comparisons.

Some of the test-cases are shown below:

1. Input: dirtbike AND terrain AND bugs
   Output:
   The no. of comparisons done is 361
   The no. of docs retried is : 2
   The list of doc ids is:
   rec.motorcycles\103122
   rec.motorcycles\103229

2. Input : hansen AND planetorium OR sky AND telescope
   Output:
   The no. of comparisons done is 460
   The no. of docs retried is : 18
   The list of doc ids is:
   sci.space\59871
   sci.space\59873
   sci.space\59874
   sci.space\59905
   sci.space\59909
   sci.space\61140
   sci.space\61219
   sci.space\61237
   sci.space\61253
   sci.space\61435
   sci.space\61476
   sci.space\62411
   sci.space\62413
   sci.space\62429
   sci.space\62453
   sci.space\62455
   sci.space\62478
   sci.space\62709

3. Input: independence AND cryptography AND NOT mechanism
   Output:
   The no. of comparisons done is 976

The no. of docs retried is : 1
The list of doc ids is:
sci.crypt\15255

Question 2: Some sample test cases are shown below:

1. **Input**: interface is frustrating

Output:

After pre-processing the query is ['interface', 'is', 'frustrating']

The no. of documents where phrase found is 1

The list of documents where phrase found is :

comp.graphics\37914

2. **Input**: wanted to take the whole station back with me in my flight
Output :
After pre-processing the query is ['wanted', 'to', 'take', 'the', 'whole', 'station', 'back', 'with', 'me', 'in', 'my', 'flight']
The no. of documents where phrase found is 1
The list of documents where phrase found is :
rec.motorcycles\103131