Name:- Parth Patel

1. Investigate two search engines (e.g., Google and Bing), and figure out what tricks they are using by explaining your testing method. (15%)  
   (a) Do both search engines use stemming?  
   To figure out if google or bing uses stemming or not, we’ll be running list of words that have same stem. For this we have taken stem words   
   1) prim:- prime, primary, primo  
   2) aqua:- aquarium, aquatic, aquaplane

Following are the such results for google and bing on those following words,

Google:

Graphical user interface, text, application

Description automatically generated Graphical user interface, text, application

Description automatically generated

Graphical user interface, application, website

Description automatically generated Graphical user interface

Description automatically generated   
  
Graphical user interface, text, application

Description automatically generated Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated Text

Description automatically generated  
  
Bing:   
  
Graphical user interface, text, application, email

Description automatically generated Graphical user interface, text, application, email

Description automatically generated  
   
Graphical user interface, text, application, email

Description automatically generated Graphical user interface, application, website

Description automatically generated

Graphical user interface, website

Description automatically generated Graphical user interface, text, application, email

Description automatically generated  
  
Graphical user interface, text, application, email

Description automatically generated Graphical user interface, text, application, email

Description automatically generated  
As we can see, both Google and bing provided different results for similar stem words, searched results have been always related to the word and not it’s stem. Eg:- Aquarium showed nearby aquariums and aquaplane showed what aquaplane means and videos explaining it. Hence, we can conclude that Google and bing don’t use stemming in their search engines and always search documents based on the words  
  
(b) Do both search engines filter stop words?   
  
To investigate this, we’ll provide a query which includes stop words and one without stop words and compare the results. For this we’ll be using following queries:

1. Query 1: cricket and football
2. Query 2: cricket football
3. Query 3: Stock price of Amazon and Apple
4. Query 4: Stock price Amazon Apple  
     
   Google  
   Query 1: Query 2:  
   Graphical user interface, text

   Description automatically generated Graphical user interface, website

   Description automatically generated  
     
   Query 3: Query 4:  
   Text

   Description automatically generated Graphical user interface, text

   Description automatically generated  
     
   As seen, Google provides the same search results with and without the stop words, (Query 1 and 2 are same, so are 3 and 4), hence we can conclude that Google filters stop words.

Bing

Query 1: Query 2:   
Graphical user interface, text, application, email

Description automatically generated Graphical user interface, website

Description automatically generated  
  
Query 3: Query 4:  
Graphical user interface, website

Description automatically generated Graphical user interface, application

Description automatically generated  
  
As seen Query 1 and 2 were different, as adding “and” made a difference in search as first query shows difference between cricket and football while other showed news about cricket football  
Similarly in Query 3 and 4, you can see News for Query 3 included stop words while news for Query 4 didn’t include them and the news were different based on the stop words.  
Hence, we can conclude that Bing doesn’t filter the stop words.

1. For a particular search query, your IR system returns 14 relevant documents and 16 irrelevant documents. There are a total of 80 relevant documents in the collection. (10%)  
   (a) What is the precision of the system on this search?  
     
   Precision is given as:  
   Precision : Fraction of retrieved docs that are relevant to user’s information need. Since, the query returned total of (16+14 =30) documents and only 14 of them were relevant, the precision of the system on this search is:

= 14/ 30

= 0.467

(b) What is the recall of the system on this search?

Recall is given as:

Recall : Fraction of relevant docs in collection that are retrieved. Since, total relevant documents in corpus were 80 and our query returned 14 of them, the recall of the system on this search is:

= 14/80

= 0.175