

Information Retrieval

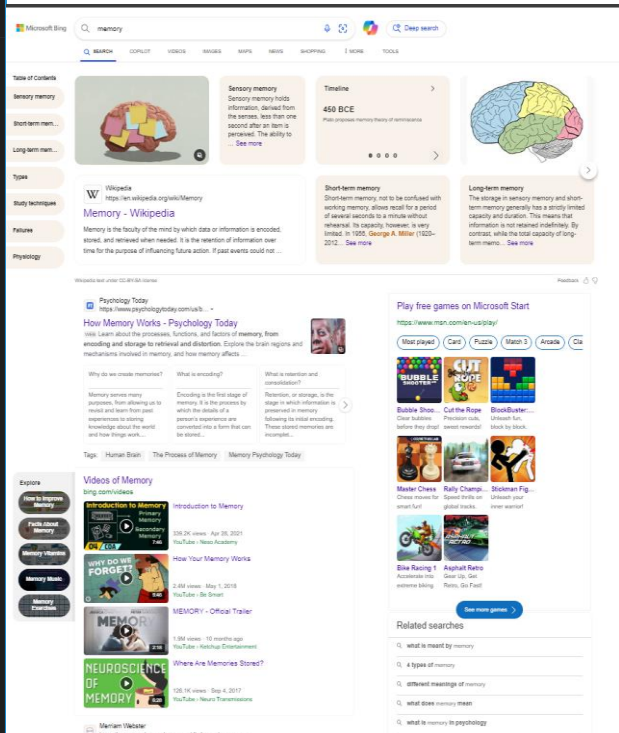
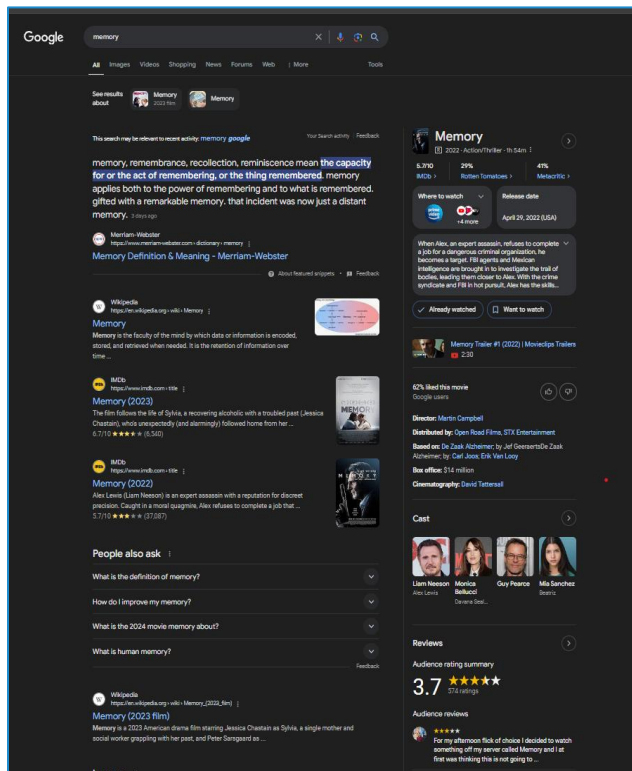
Assignment-2

1. Investigate two search engines (e.g., Google and Bing), and figure out what tricks they are using by

explaining your testing method.

(a) Do both search engines use stemming?

For both the search engines Google and Bing, we will test a few queries and check what results do we observe. So, for stemming we will run the query “memory”:



Now, we'll run the query "memorial":

Google search results for the query "memorial". The results are filtered for "Worcester, MA 01609". The top result is a dictionary entry for "me-mo-ri-al" (noun), defining it as "something, especially a structure, established to remind people of a person or event." Below the dictionary entry, there are two listings for "UMass Memorial Health" and "UMass Memorial Medical Center in Worcester, MA". At the bottom, there is a map showing the location of the UMass Memorial Medical Center in Worcester, MA.

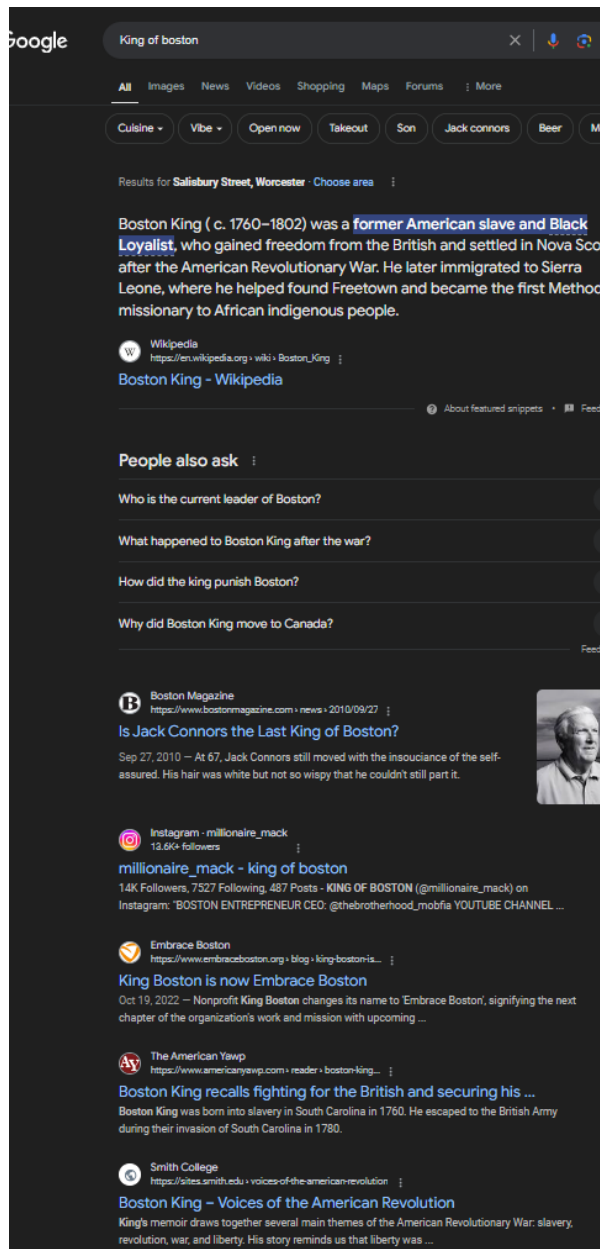
Microsoft Bing search results for the query "memorial". The results show "About 1,830,000 results". The top result is a Merriam-Webster entry for "Memorial", defining it as "Learn the meaning, synonyms, examples, and history of the word memorial as an adjective and a noun. Find out how to use memorial in a sentence to commemorate a person or ...". Below the dictionary entry, there is a section for "Videos of Memorial" with several video thumbnails. At the bottom, there is a section for "Short videos of memorial" with several video thumbnails. On the right side, there is a sidebar with a "Massachusetts Memorial" listing and a "Reviews" section.

So, as we observed for both search engines:

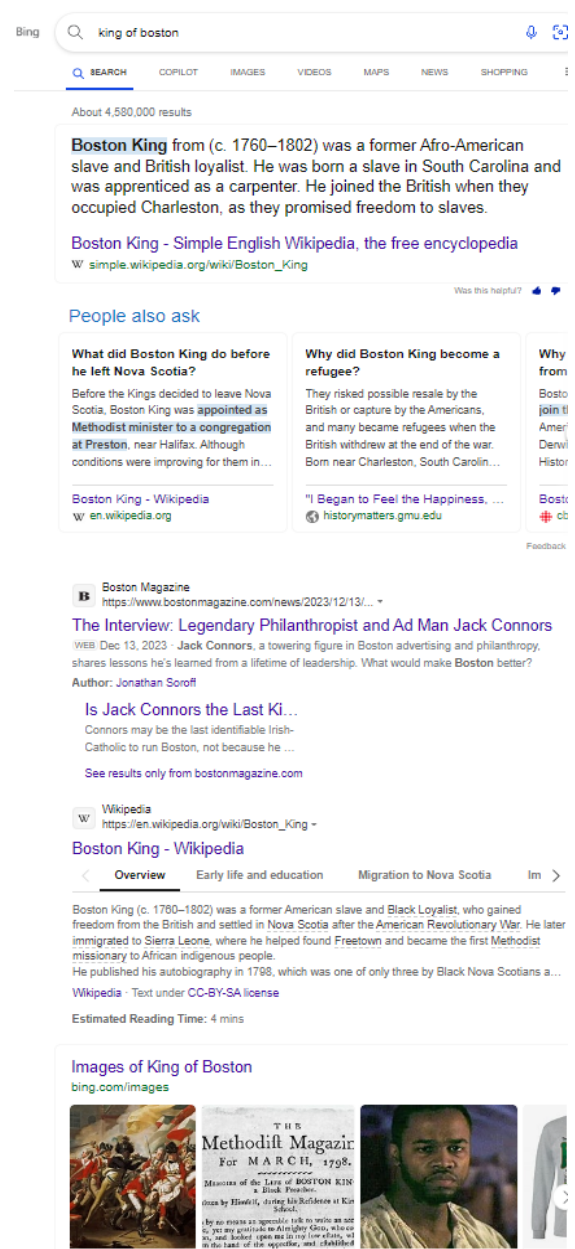
- If stemming is used, we'll get the similar results for both the queries, memory and memorial
- If stemming is not used, then we would get different outcomes which would focus on the main query term.

(b) Do both search engines filter stop words?

We will test a few queries here as well and check what results do, we observe. So, for filtering stop words we will run the query “King of Boston”:

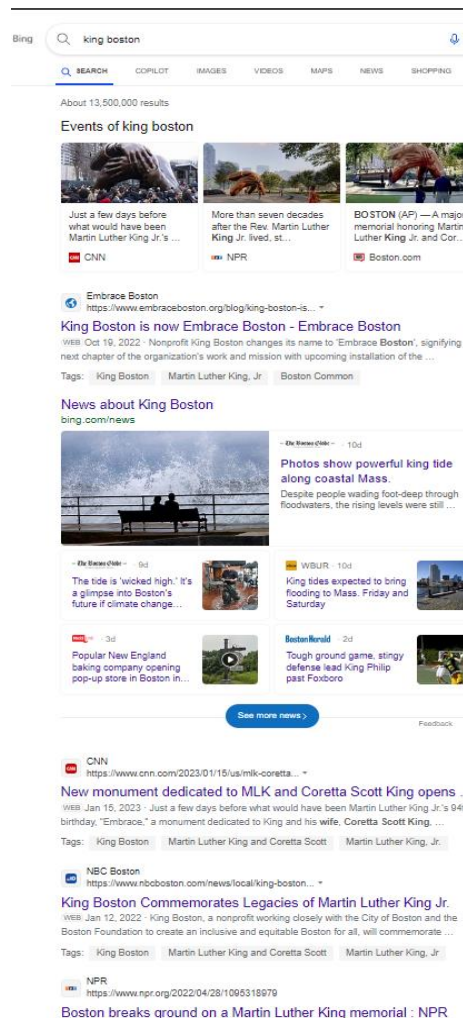
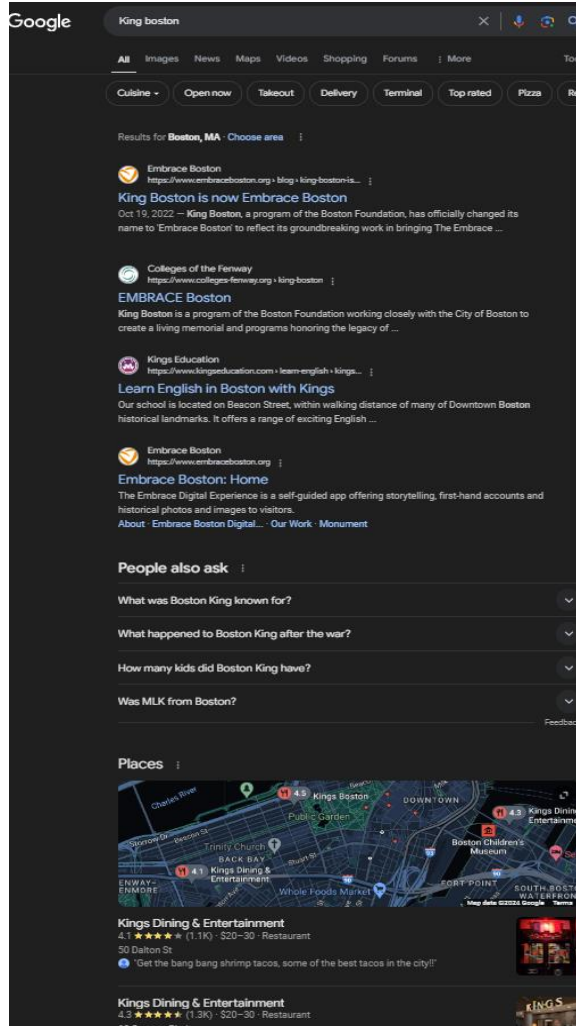


Google search results for "King of boston". The search bar shows "King of boston" with a microphone icon. Below the search bar are tabs for "All", "Images", "News", "Videos", "Shopping", "Maps", "Forums", and "More". A filter bar shows "Cuisine", "Vibe", "Open now", "Takeout", "Son", "Jack connors", "Beer", and "Me". The main results section shows a snippet from Wikipedia: "Boston King (c. 1760–1802) was a former American slave and Black Loyalist, who gained freedom from the British and settled in Nova Scotia after the American Revolutionary War. He later immigrated to Sierra Leone, where he helped found Freetown and became the first Methodist missionary to African indigenous people." Below this is a "People also ask" section with questions like "Who is the current leader of Boston?", "What happened to Boston King after the war?", "How did the king punish Boston?", and "Why did Boston King move to Canada?". At the bottom, there are several articles and social media posts, including one from Boston Magazine titled "Is Jack Connors the Last King of Boston?" and one from Instagram titled "millionaire_mack - king of boston".



Bing search results for "king of boston". The search bar shows "king of boston" with a microphone icon. Below the search bar are tabs for "SEARCH", "COPILOT", "IMAGES", "VIDEOS", "MAPS", "NEWS", and "SHOPPING". The main results section shows a snippet from Wikipedia: "Boston King from (c. 1760–1802) was a former Afro-American slave and British loyalist. He was born a slave in South Carolina and was apprenticed as a carpenter. He joined the British when they occupied Charleston, as they promised freedom to slaves." Below this is a "People also ask" section with questions like "What did Boston King do before he left Nova Scotia?", "Why did Boston King become a refugee?", and "Why from Boston?". At the bottom, there are several articles and social media posts, including one from Boston Magazine titled "The Interview: Legendary Philanthropist and Ad Man Jack Connors" and one from Wikipedia titled "Boston King - Wikipedia".

Now, we'll run the query "King Boston":



So, as we observed for both search engines:

- If stop words are filtered, we'll get the similar results for both the queries, which tells us that the stop words are ignored
- If stop words are not ignored, then we would get different outcomes, and sometimes the stop words would play an important role for finding documents based on the given query term.

Finally, we observe that:

- a. Both **Google** and **Bing** are expected to use stemming to some degree, but the results differ as per their behavior in stemming algorithms.
- b. Both search engines are likely to filter stop words, but don't have that much difference if the query is with or without stop words.

2. 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.

Ans.

Here, we have:

14 relevant documents retrieved

16 irrelevant documents retrieved

Total number of retrieved documents = $14 + 16 = 30$

Total number of relevant documents = 80

(a) What is the precision of the system on this search?

Precision is number of documents retrieved which are relevant to the given search query, which is also calculated as accuracy

$$\begin{aligned}\text{Precision} &= \frac{\text{Number of relevant documents retrieved}}{\text{Total number of retrieved documents}} \\ &= \frac{14}{14+16} = \frac{14}{30} = 0.4667 = 46.67\%\end{aligned}$$

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

Recall is a fraction of the number of relevant documents which were retrieved by the total number of relevant documents present in the collection.

$$\begin{aligned}\text{Recall} &= \frac{\text{Number of relevant documents retrieved}}{\text{Total number of relevant documents}} \\ &= \frac{14}{80} = 0.175 = 17.5\%\end{aligned}$$