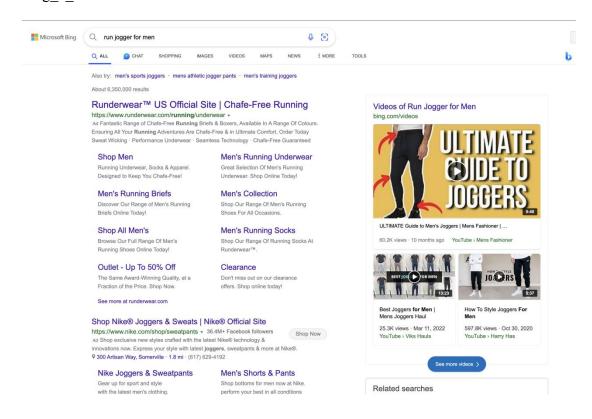
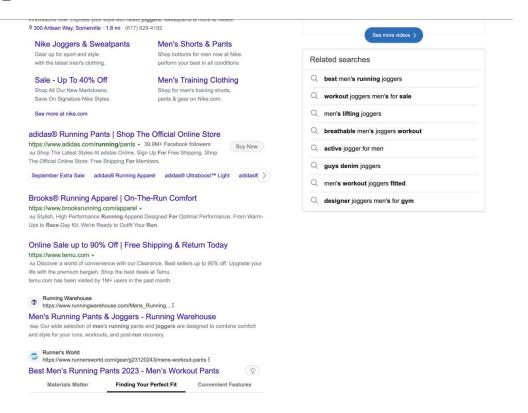
- 1. Investigate two search engines (e.g., Google and Bing), and figure out what tricks they are using by explaining your testing method. (15%)
- (a) Do both search engines use stemming?
- (b) Do both search engines filter stop words?

Yes I believe that Google & Bing both search engines uses stemming and filters stop words while fetching the results for the user's query. I've tried two examples on both the search engines and the ranking in is different in Google's results but stays somewhat same in Bing.

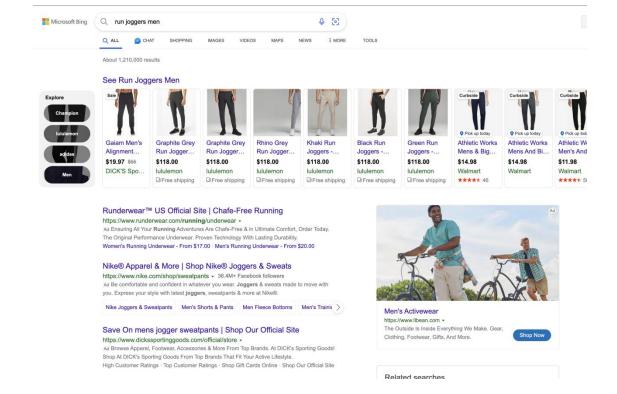
bing a 1



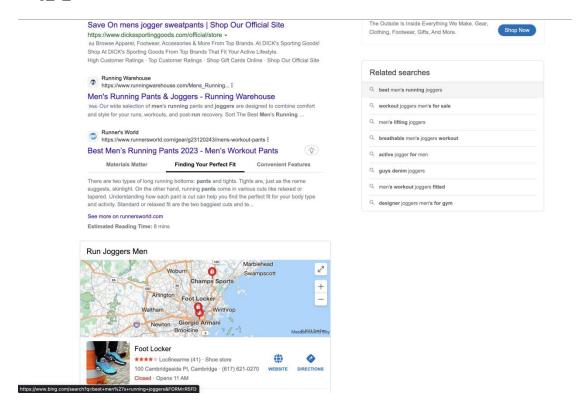
bing_a_2



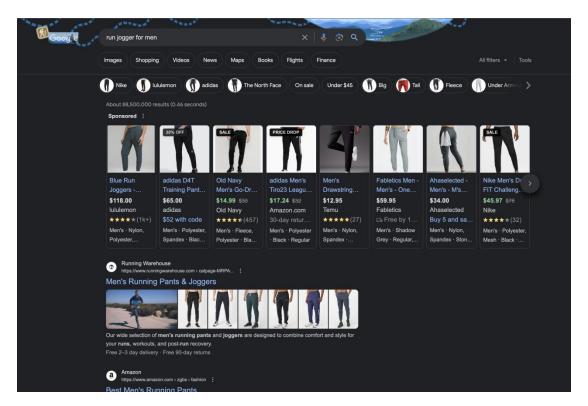
bing_b_1



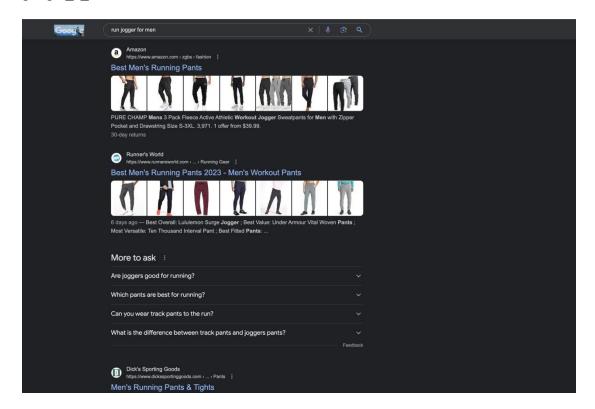
bing_b_2



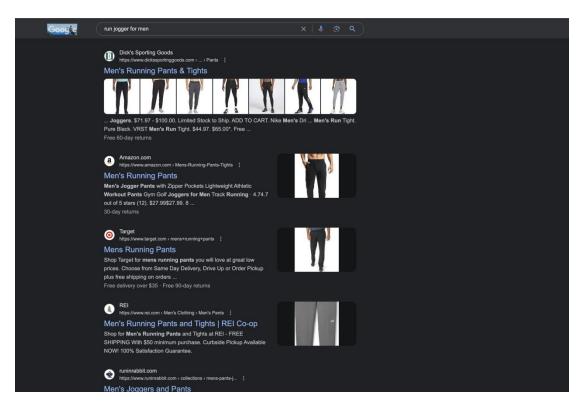
google_a_1



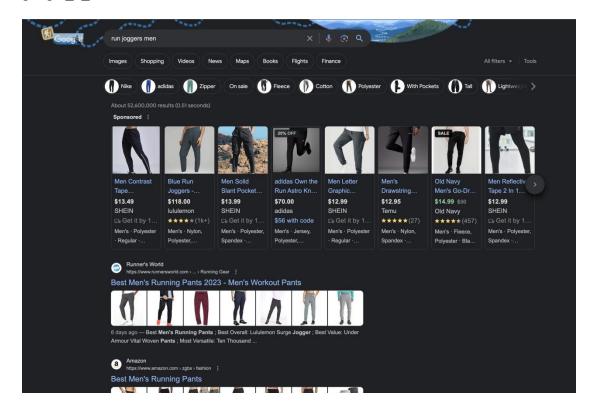
google_a_2



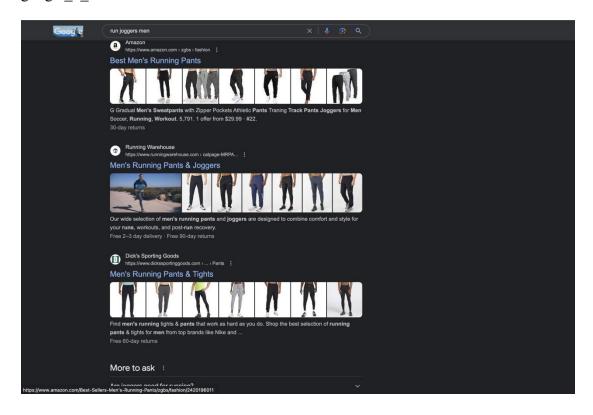
google_a_3



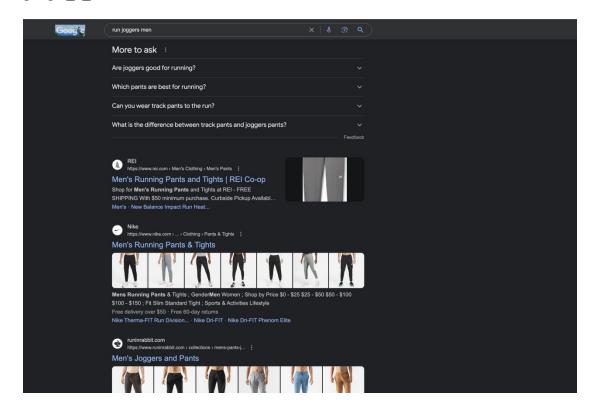
google_b_1



google_b_2



google_b_3



- 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. (10%)
- (a) What is the precision of the system on this search?
- (b) What is the recall of the system on this search?
- (a) Precision = (14 relevant documents) / (14 relevant documents + 16 irrelevant documents) = $14 / 30 \approx 0.4667$ or 46.67%
- (b) Recall = (14 relevant documents) / (total relevant documents in the collection) = $14/80 \approx 0.175$ or 17.5%

So, the precision of the system for this search is approximately 46.67%, and the recall is approximately 17.5%.