

In-memory Search engine

Your organization has started a new tech blog with interesting tech stories and you're responsible for designing and implementing an in-memory search engine, supporting the search functionality on the blog content.

Requirements **TO SHARE WITH THE CANDIDATE**

- A. It should be possible to create a dataset in the search engine.
- B. It should be possible to insert documents into a given dataset. Each document is simply a piece of text.
- C. It should be possible to search through documents for a search pattern in a given dataset.
- D. It should be possible to order the search results

The following is **FOR THE INTERVIEWER ONLY**

It should support the following.

(The text in bold is the primary problem statement. The rest of it can go to requirement analysis for SDE 2-3. Refer: [SDE 1-3 Coding round expectation matrix](#).)

1. **It should be possible to create a dataset in the search engine.**
2. **It should be possible to insert documents into a given dataset. Each document is simply a piece of text.**
3. **It should be possible to search through documents for a search pattern in a given dataset. A search pattern could have one or more words. When there are more words, all the words in the search pattern should be present anywhere in the document to match, irrespective of their order.** Assume that search is case-insensitive.
4. It should be possible to configure each dataset on how to order the search results. (Note for the interviewer: For example, one may want to configure a dataset to give higher rank to the documents with more number of occurrences of search term or latest)
5. It should be possible to print search results of a search.
6. It should be possible to print info of a given dataset - configuration and the document count.
7. It should be possible to delete a dataset. This should delete configuration of it along with all the documents in the dataset.

Add-ons:

8. It should be possible to configure each dataset on how to split the documents too, to be able to match with the search terms. (Note for the interviewer: For example, we may want to split the document on any punctuation mark instead of whitespace.)

Examples

Search term: apple

Doc1: apple is a fruit

Doc2: apple apple come on.

Doc3: oranges are sour

Doc4: apple is sweet

Search results (number of occurrences of search term ranker)

[Doc2, Doc1, Doc4] or [Doc2, Doc4, Doc1]

Search results (recency ranker)

Doc 4, Doc2, Doc1