

ITM Assignment

(Arnab Baranwal
AB -- 2820208)

Q1. What are the challenges faced by search engines?

A search engine is a software system that is designed to carry out web searches. They search the world wide web in a systematic way for particular information specified in a textual web search query.

The reality is that search engines are far from perfect. They face significant challenges trying to decipher our intentions and find the pages that match with our queries.

Here are some of the problems-

1. Repetitive title tags =>

The most common is the number of repetitive title tags used. This is something Google (search engine) heavily penalizes for and something that is easily fixed.

2. Unclean URLs =>

Search engine spiders have to be very efficient, so, they are leery of anything that seems like a risk to them, even though it may not be. Dynamically generated pages present a risk to them because the spiders could get caught up in an infinite loop within the site. URLs having long, complex query strings will have a harder time getting indexed with everything else being equal, than a shorter URL with no special characters.

3. Spam => Search engines identify relevant pages by means of 'quality signals' or metrics that can be deduced from web pages by automated means. The challenge to search engines is that once black hat SEOs identify those signals, they can start to fake

(282020)
them. For humans it is easy to spot spam, but for computers it is much harder.

4. Polysemy \Rightarrow There are words that change their meaning depending on the context in which they are used.
e.g. if you do a search for 'wood' you might see pages that are talking about pieces of a tree, or you might be talking about the geographical area that has many trees. Without the right context, it is hard for a human to tell. Imagine how hard it is for the search engine!

5. Poor queries \Rightarrow Many searchers don't know how to express what they want in the real world, and are even worse attempting to ask a search engine. Worse yet, they misspell words, making the problem more 'interesting' for search engines.

6. Synonymy \Rightarrow This is another challenge. There are words that have the same meaning, like 'car' and 'automobile'. When you do a search you would like to get pages that contain your exact words, and pages that contain other words that mean the same thing, as long as they are relevant to your search. These details don't pass through the minds of most searchers, but search engines have a hard time because of it.

7. Natural language searches \Rightarrow

A MySQL database engine can precisely return all the relevant records given a query 'select first last from employee where last = "Smith"'. There is a formal syntax and no ambiguity. A search engine on the other hand, receives 'who has smith as last name in Chicago' or 'smith last name Chicago'.

(2820208)

query is in natural language -- our language. There are many different ways to say the same thing i.e. there is context, there are human idiosyncrasies, and so on. The search component of a search engine must disambiguate the query and translate it into a more formal manner before looking it up in the index.

8. It can't be it all.

Indexing all of the content on the web regularly is more than just a daunting task; it's impossible. Google (a search engine) has indexed billions of webpages, but it has not and likely will index every single piece of content on the web. That's a result of the robustness of the web's content offering, and Google also admits that it cannot all elements of websites effectively.

Whether text is easy to crawl and index + while images, videos and certain HTML elements are not so easily crawl and indexed.

Q. What is meta search engine? How it works and also explain characteristics of meta search engines?

The Metasearch Engine is a search engine that combines the results of various search engines into one and gives one result. It can also be stated as an online information retrieval tool.

In simple terms, a metasearch engine takes the query you've entered and gathers results from multiple search engines online, such as Google, Bing, Yahoo, and more. They aggregate the results for you so you can choose the best information from the search results provided.

(2820208)

Any search engines receives several queries every day. The meta search engines run the queries on multiple other search engines and in turn reflect the results in the form of the summarization of such sites.

Working

So now back to how a meta search engine works. Meta search engines don't have a repository or index of their own, they take advantage of indices created by other search engines. In fact, they present this as their strong point. A typical meta search engine pulls off the results from a number of search engines, say Google and Bing, and then apply their own algorithms in some cases to re-order the results.

The obvious thought that comes to mind is that meta search engines get their results from multiple search engines and since there is better than one the results must also be better. The truth however is different, in fact let me put it this way, it is matter of preference.

~~Any search~~ engine does not create a database of A meta search it creates a federal database that itself rather an integration of the databases of various other search engines.

Characteristics of Meta-Search engines

Architecture of Ranking ⇒

1. A meta search engine develops its own algorithm

it eliminates duplicate results and calculates and ranking of the sites. This is because it understands that the websites which are highly ranked on major sites are more relevant and would thereby provide better results.

2. Fusion \rightarrow Fusion is used to create better and efficient results. Fusion is divided into collection Fusion and Data Fusion. The collection Fusion deals with search engines that contain unrelated data. The data sources are then ranked based on their content and the likelihood of providing relevant data. This is then recorded in a list. The Data Fusion deals with the search engines that have indexes for common data sites. The initial ranks of the data are compared with the original ranks. A process of Normalization is applied using techniques such as the combSum algorithm.

3. The meta search engines search for all the search engines at once hence individual search is unnecessary.

4. Searching for the white pages and yellow pages are allowed in the metasearch engine.

5. The users can customize the metasearch engine according to their preferences.

6. It reduces the workload of the users by allowing comprehensive results in every little time skipping the time and toil of searching for results in the individual sites.

7. Metasearch engine is not capable of parsing.
8. The metasearch engine is not capable of fully translating the query syntax.

9. It comes free of cost and they usually prioritize the pay per click types of links before the more relevant search results.

10. The Metasearch engine provide a means of hiding the IP addresses from the searches. They provides a lot of security to the user.

This is the ~~same~~ reason why the search government uses Bount (metasearch engine) for its operations.