All About The Search Engine of Leetcode

report by Aiqi Jiang, CS6200, A4, Nov.18

My choice of search engine is Leetcode.com. Below is all information about the search engine after I study it by online investigating and query testing.

Brief Introduction of Leetcode search engine

Firstly, we care about the category of searched material in Leetcode. Leetcode website provides various coding questions for coding lovers and professional programmers. It has over 1,600 different problems now, support for around 15 programming languages. Plus, it offers an active community that is always there to help you with the solutions you come up with. [1]Users can also search for questions and comments of various topics and types, such as interview questions, interview experience, career and so on. As a powerful search engine Leetcode keeps growing and enlarging itself to meet more demands for all areas of programmers' career life.

The search engine is of great interest to me because it is the most helpful website I used last year to solidify my algorithm and coding techniques. Besides, The search engine is very user-friendly, it provides various searching strategies to locate and screen out questions users want most. The whole search engine consists of several sections: "Explore", "Problems", "Mock", "Contest", and "Discuss". They come together to fulfill all needs for users to prepare for their school and career life.

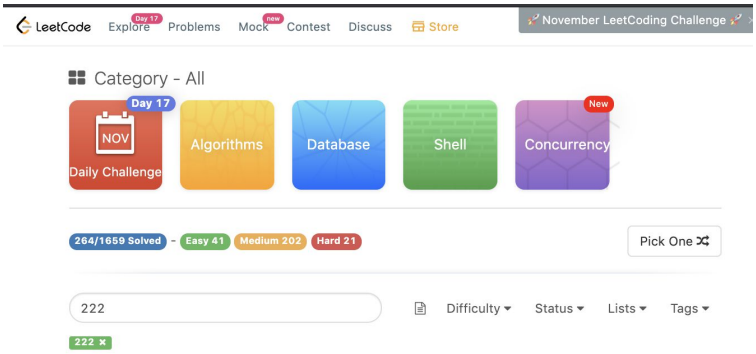

Figure1: different sections in the search engine

Search types and ranking features

This engine supports many kinds of searches. Besides typical coding problems, user-posted interview questions, interview experience, compensation information, topics of career, general discussion and even feedback for the website are also searchable in the engine. For example, in the career section, after I query "facebook amazon" and want to have a comparison of the two company, the engine quickly display some top results regarding the comparison of different facets such as leetcode problems, work choice, interview preparation, etc.; in the Compensation section, after I query "microsoft intern

[1] Jordan Bishop (2020, Oct 20).*What is Leetcode?*. Retrieved from https://www.quora.com/What-is-Leetcode

seattle", top results will show the compensation information of Microsoft internship position in Seattle posted by other users.

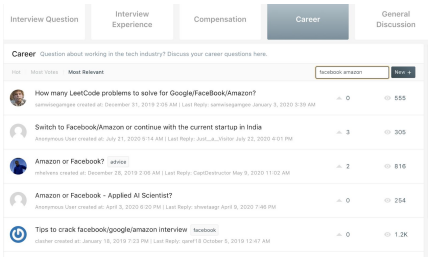

Figure2: results of querying "facebook amazon"

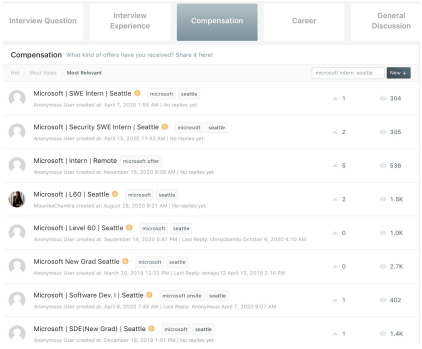

Figure3: results of querying "microsoft intern seattle"

As for the ranking strategy, Leetcode also supports various types.

1. By problems' difficulty(totally 3 levels: easy, medium and hard)
   example link: https://leetcode.com/problemset/all/?difficulty=Easy
2. By problems' status(totally 3 status: todo, solved, and attempted)
   example link: https://leetcode.com/problemset/all/?status=Todo
3. By problems' tags
   example link: https://leetcode.com/problemset/all/?search=array

Plus, there are other ranking types which doesn't show difference in url: problem ID, Title, Solution, Acceptance, Frequency; in the part of discussion, ranking types also includes:

1. Hot
   example link:
   https://leetcode.com/discuss/interview-question?currentPage=1&orderBy=hot&query=microsoft
2. Most Votes
   example link:
   https://leetcode.com/discuss/interview-question?currentPage=1&orderBy=most_votes&query=microsoft
3. Most relevant
   example link:
   https://leetcode.com/discuss/interview-question?currentPage=1&orderBy=most_votes&query=microsoft

How Leetcode engine works

In terms of the source of data Leetcode got materials from, real interview questions from tech companies are the majority of them, and the data is not available on the web already that can be crawled. [2] These data usually come from IT companies such as Google, Facebook, Amazon and people having experience of company interviews. Besides, there are many posts from online users which group into a large community. So Leetcode gets data from both companies and individuals.

There are many important features of the data, such as description, examples, related topics and test cases of each question. This type is special and specific in the components of each question such as related topics it belongs to, examples of input and output and so on.

Moreover, there are many major ranking features that it seems to use. For example, in part of interview questions, Leetcode seems to use the number of votes, the level of hotness as ranking features. People want to make them a feature.

Leetcode seems to use hot most in the section of interview questions, votes most and relevant as major ranking features. They are special for category since most of the questions are posted by online users, other users can vote for their favourite answers. So as time goes by, the quality of each post will be fairly calculated by these ranking features for people to find answers fitting their demands much faster.

Many people wonder how the system seems to rank results for the given query. Although there's no information available about this, I have an assumption of the work based on my testing of different queries in the engine: before the searching begins, Leetcode has already operated index construction on all questions in their database. Then Leetcode will separate the query into tokens, and for each token the search engine will calculate a weighted tf-idf vector for it. After evaluating the weighted tf-idf vector for each question text, Leetcode calculates score for each question as the basis of ranking.

Technical problems and ways to solve them

When users need to search out questions for preparation of certain company interviews, they can either directly type in the company name or input the company's name in a separate search box named "Tags". There are two technical problems the creators must have solved, one of them is: how to search out all interview questions accurately and order them by relevance according to the description words users type in?

---

[2] Hao Zhang (2016, Jan 4).*Where does Leetcode get its questions*. Retrieved from https://www.quora.com/Where-does-Leetcode-get-its-questions

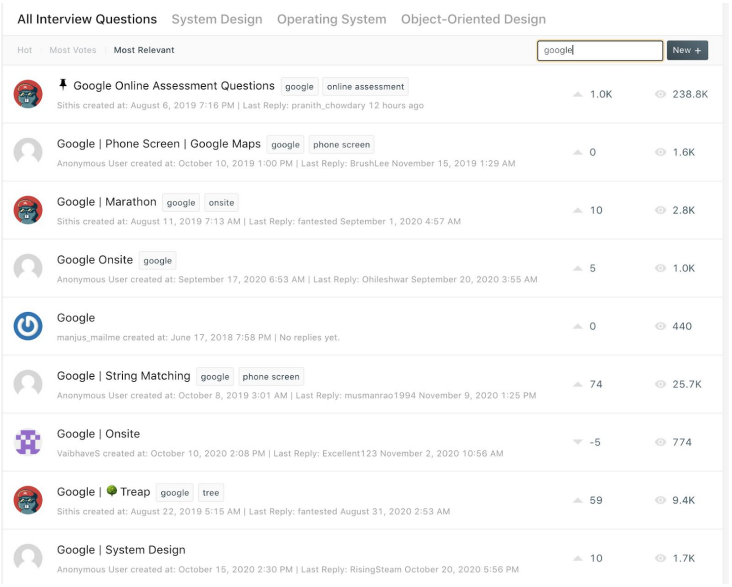

figure4: results after querying “google”

Besides, I notice that during the process of typing “Google” in the “Tags” section character by character, the search engine keeps updating the suggested results below. So there comes another technical problem: how to enable the function of updating the suggested results below?

figure5: function of tags

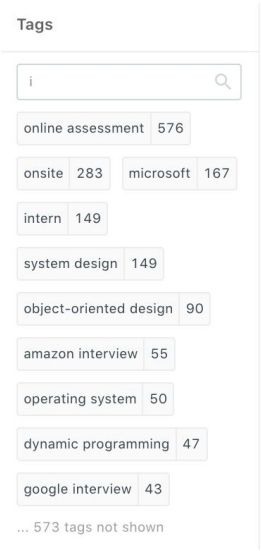

figure6: function of updating tags results

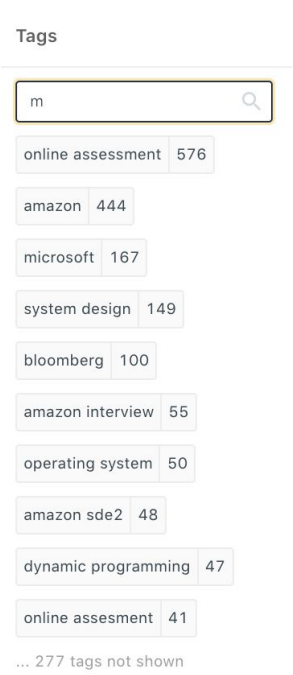

figure7: function of updating tags results

Here are my ideas about how to solve them:

For the first one, I think they will first search out all candidates with tags including the company's name or having the company's name in posts' titles or full text contents. Then use tf-idf weighting to calculate ranking scores and sort them from highest to lowest. I also guess that the weight of tags should be bigger than the occurrence of the company's name in text, since tags of each result are manually added by online posters, so they are more accurate and important.

For the second problem, I tested different queries such as "facebook", "google" and "intuit" and find that when I firstly input "i" for "intuit", the top results display tags containing the character "i" ordering by the number of posts shown at the right of each tag, but not tags starting with "i". This situation remains until I input "intu" and there finally lists "intuit". From

the tests, I see how creators make the search engine keep updating suggested results while we are typing queries character by character: every time I add a character, the search engine updates the results by listing out all tags containing the current query ordering by post amounts until we finish typing.

Strength and limitations of Leetcode search engine

Leetcode is best for clustering and topics/tags classification. As for topics/tags classification, we can see that every problem has one or more tags, which enables users to search out coding problems of certain topics very fast.

Plus, each problem has similar questions given by Leetcode. This feature is enabled by clustering's document grouping. Leetcode groups all problems into different topics, so all similar questions listed below the problem have at least one same topic or label with it. So it is convenient for users to continue to do practice on similar questions.

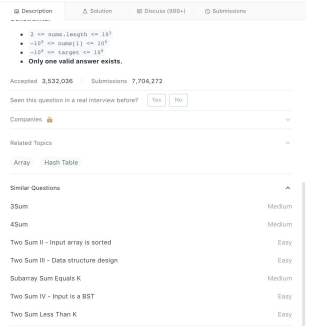

figure8: clustering strategy used for showing similar questions

Leetcode also has limitations, especially when implementing query searching. For example, in the section of "Contest", users can't query out certain contests with their id numbers. They can only browse every contest page by page. As for this problem, I think Leetcode can solve it by adding one searching feature of querying out the problem with its id number, for example "215".

Besides, in the section of "interview questions" , although we can filter out questions with company names, further filtering by different positions is not allowed so it is hard to justify whether the post is for interns, junior software engineers or senior positions. Here's what I think ways to solve this problem: When someone makes a post for the interview question section, Leetcode search engine can ask him/her to choose lists of position choice: intern, junior software engineer, senior software engineer, etc, so that users can quickly search out posts of interview experience conforming to their positions.

Suggestions for improving Leetcode search engine

Lastly, I have some suggestions for Leetcode search engine to improve its search performance. Firstly, I think Leetcode can optimize the tagging strategies for users. For example, in the part of discussion, users are asked to tag topics when adding new posts. And I think Leetcode can better the process of labeling by adding categories of labels so that other users can clearly search out posts meeting their requirements. For example, in the section of career and general discussion, the categories of labels can be separated into company name, job positions, types of discussion(interview tips, salaries, work-life balance, etc.) Also Leetcode can add drop-down menu listing suggestions of existing labels below

these categories so that users can better refine categories of labels and other users are able to quickly find out their ideal answers.

Moreover, I suggest that Leetcode add more functions to merge social properties into its platform. For example, adding a social forum to enhance connections between users and their friends. In this forum, users are able to display their progress of solving coding problems or communicating with their friends about interview experience together. In this way people can be encouraged to keep progressing further.