Useful References

Here we want to evaluate how to prioritize the links/references that we would display to the user when they query for something. The idea is to prioritize and display the Stack Overflow questions that are most relevant to the query entered on the top.

Here are some of the criterions that we can evaluate to assign a score to how relevant a particular Stack Overflow question link is to the query entered:

1. **Tag Match Score:** The number of tags that our query matches to in a question. We can compare the user's query word by word to the tags of a question to determine how relevant the two are. A question with more tags matching to the query is more relevant to the query than others.

2. **Views**: The number of views a question has gotten. The more the number, the more was that particular question looked at or referred to by people.

3. **Accepted Answer:** It's a lot convenient to show questions with an accepted answer on top because the user is then assured that his query has been answered.

4. **Number of answers**: The number of answers a question has can reflect diverse opinions/explanations on/to the question. The more answers a question has, the more different approaches to answering the question have been presented, giving the user a variety of approaches to look at.

   Of course, this is not always the case. In some of the questions, a lot of the answers posted to a question could be the exact same approach/solution, thereby making those answers just duplicates of each other. So, this factor should be given the least priority so that we don't spam the user with questions that have multiple but same answers.

5. **Votes**: StackOverflow accepts votes on both questions and answers. We can evaluate each to figure how useful the question is to the query.

   -> Question Votes: The more the votes, the more in demand was an answer to the question. It indicates how popular the question was in the topic.

   -> Answer Votes: The number of answers plus the votes on each answer gives us an idea of how many people found the answer (and thus the question too) useful. Since questions have different number of answers, we would have to normalize this factor to give us a good indication of which question has more useful answers.

   We could average the count out as (Total Number of Votes across all answers / Total number of answers) to give us a metric to compare two different questions with.

   So if we have one question with 2 answers, 1 voted 36 times and another voted 100 times and another question with 7 answers voted 2, 50, 5, 150, 1, 3, 30. If we

average the count out, the first question comes up with a score of 68 and the second question comes up with a score of 34.42, therefore implying that the first question is more useful, even though it has lesser number of answers and lesser number of total votes.

We can prioritize the questions we want to display using the above factors and generate a score out of them. The higher the score, the higher up the question is in the order.

The order in which the factors contribute to the usefulness of a link would be (highest to lowest):

F1. Number of tags matched (TM)
F2. Number of votes on a question (QV)
F3. If the question has an accepted answer (AA)
F4. Averaged votes on answers (AAV)
F5. Views (V)
F6. Number of answers (NA)

We can assign each factor a weight and then calculate a total score out for the link as:

Score = (5 * TM) + (4 * QV) + (3 * AA) + (AAV) + (V) + (0.5 * NA)

(We don't want the number of answers to dominate our score, that is why we can assign it a weight of 0.5).