Q1: How did you calculate the similarity of the queries and why did you choose this method? What are its limitations?

Model Used and why:
I used TF-IDF "Term Frequency-Inverse Document Frequency" NLP Model to approach the problem. This model provided a simple and effective way to represent text data as numerical vectors that can be used for machine learning algorithms. By converting text data into numerical vectors, it becomes possible to apply a wide range of machine learning techniques to solve the Query Search Engine Problem.

Limitation:
Word order: The TF-IDF model does not take into account the order in which words appear in a document, which can be important for certain types of text analysis, such as sentiment analysis or text generation.

Synonymy and polysemy: The TF-IDF model treats each word as a separate entity, which can lead to difficulties with words that have multiple meanings or synonyms. This can result in loss of context and potentially inaccurate representation of the meaning of the text.

Overemphasis on rare words: The TF-IDF model gives high weight to rare words, which can be problematic if the rare words are not relevant to the topic or if they are misspelled or noisy.

Q2. How would you quantify how well the similarity service is doing? How could it be improved?

To improve the results of a search query engine even further, there are several strategies that can be used in conjunction with the TF-IDF model:

Word embeddings: Word embeddings are a way of representing words as numerical vectors that take into account both their frequency and their context in a document or corpus. This can help address some of the limitations of the TF-IDF model, such as the lack of consideration for word order and synonymy.

Deep learning models: Deep learning models, such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs), can be used to learn more complex and nuanced representations of text data, which can improve the accuracy of search query engines.

Domain-specific dictionaries: To improve the accuracy of the TF-IDF model, it can be useful to create domain-specific dictionaries that contain relevant keywords and phrases for the topic being analyzed. This can help to ensure that the TF-IDF model is focusing on the most relevant terms and avoiding noise.

Overall, while the TF-IDF model is a powerful tool for natural language processing, it is important to recognize its limitations and to use complementary strategies, such as word embeddings or deep learning models, to improve the accuracy of search query engines even further.