CSC 820
HW 11 Report
Andrew Dahlstrom
4/29/24

Evaluation of Term Frequency and Term Frequency-Inverse Document Frequency

Term Frequency (TF) measures the frequency of a query's terms within each document, potentially favoring longer documents or those where query terms frequently occur. This can lead to higher rankings for documents that might not necessarily be more relevant but simply contain more words or more instances of some common terms. Such an approach might skew results towards documents that are less relevant, especially if context and style across documents isn't considered.

On the other hand, Term Frequency-Inverse Document Frequency (TF-IDF) enhances the document ranking process by considering not only the frequency of query terms within a document but also their rarity or significance across all documents. This results in higher rankings for documents that contain terms which are both relevant to the query and distinctive to the corpus of documents. TF-IDF is therefore better for situations where the uniqueness of document content is key, such as in academic research papers. It effectively prevents common terms from overshadowing more relevant terms, thus prioritizing documents that offer more unique and relevant information. In scenarios where the frequency of terms directly correlates with relevance, such as certain legal or technical texts, TF might still be appropriate.