

Abstract

This paper presents a framework for aggregating and retrieving relevant maize information using Term Frequency Inverse Document Frequency and Term Proximity. The framework aggregates information from agricultural websites and blogs through the use of RSS technology. Term Frequency Inverse Document Frequency is able to retrieve relevant documents from the aggregated RSS feeds however; the presence of a query term within a retrieved document does not necessarily imply relevance. Documents with same similarity score do not necessarily have the same level of relevance. To mitigate that problem we implement a term proximity scoring approach that will be able to improve relevance in the top-k documents returned by TF-IDF. The approach for term proximity score uses both the span-based method and pair-based method to ensure effective proximity scoring. User preference profile is based on keywords which form user query while text documents are composed of RSS description content and RSS title tag content. Stemming is applied on query and document terms for better precision. This framework will ensure maize farmers get the most relevant information from online sources.