**AIWR Paper Report**

**SRN:** PES1UG20CS109 **Section:** B

**Name:** Chetan Reddy Bandi

**Title:** **A Pseudo-relevance feedback framework combining relevance matching**

**and semantic matching for information retrieval**

Introduction:

The accurate retrieval of relevant information from a vast pool of documents is an essential task in natural language processing. Pseudo-relevance feedback is a technique that can enhance information retrieval by generating additional relevant terms for a user's query using a set of top-ranked documents. The authors of this paper introduce a novel pseudo-relevance feedback framework that blends relevance matching and semantic matching to boost the precision of information retrieval.

Methodology:

The proposed framework is composed of two main elements: relevance matching and semantic matching. The relevance matching element uses traditional term-based techniques like BM25 to determine the similarity between the query and each document in the collection. On the other hand, the semantic matching element employs pre-trained language models, such as BERT, to evaluate the similarity between the query and each document based on their semantic meaning. The top-ranked documents from the relevance matching element are then used as input to the semantic matching element, which generates additional relevant terms for the query based on their semantic similarity to the top-ranked documents. The authors then combine the original query with the generated terms using a weighted approach to re-rank the documents and produce a final list of results.

Results:

The proposed framework was assessed by the authors using TREC-COVID and Robust04, two publicly available datasets. The results indicated that the proposed framework surpassed several baseline methods in MAP and NDCG metrics. Additionally, the authors carried out an ablation study to demonstrate the effectiveness of each framework component.

Conclusion:

The authors proposed framework which utilizes a combination of relevance matching and semantic matching to enhance the accuracy of information retrieval. Through experimentation, the authors have demonstrated the effectiveness of this new pseudo-relevance feedback framework in retrieving relevant documents from vast collections. The framework has the potential to boost the performance of current information retrieval systems and can be implemented in various domains.

**Title: A dummy-based user privacy protection approach for text information retrieval**

Introduction:

In recent times, the progress of technology has facilitated the effortless access to copious amounts of information via text information retrieval systems. However, this convenience also brings to light apprehensions regarding user privacy, as the personal information of users may be utilized to offer customized search outcomes. To address this issue, the authors suggest a novel approach that employs dummy data to safeguard user privacy without compromising on the accuracy of the search results in the realm of text information retrieval.

Methodology:

In this proposed approach, dummy queries are created based on the user's original query to protect their personal information. These dummy queries are generated using terms that are similar to the original query, but do not reveal any personal information. The authors train a privacy model using machine learning techniques to differentiate between the user's original query and the dummy queries. The privacy model is then utilized to filter search results and display only the most relevant ones based on the user's original query, while excluding any results related to personal information. The authors conducted an evaluation of their approach using a real-world dataset of web search queries and compared it with k-anonymity and randomization methods. The evaluation metrics, including precision, recall, and F1-score, were used to measure the effectiveness of their approach.

Results:

The evaluation indicates that the proposed approach excels in maintaining user privacy while delivering precise search results, surpassing the performance of the baseline methods. The approach achieves high levels of precision and recall, while efficiently concealing the user's personal information from the search engine.

Conclusion:

The proposed approach by the authors employs a machine learning algorithm to generate dummy queries and train a privacy model for preserving user privacy while ensuring accurate search results in text information retrieval. The approach is versatile and has been demonstrated to be effective in maintaining user privacy without compromising the search outcomes, as shown by the experimental results. The authors suggest that this technique can be utilized across multiple domains and could enhance the privacy of current text information retrieval systems.