

Retrieving BioMedical Information: Challenges and Possibilities

Heri Ramampiaro
Department of Computer and Information Science
Norwegian University of Science and Technology (NTNU)
Trondheim, Norway

Abstract

A large amount of biomedical information is available to researchers today, and it is continuously increasing. As a result, researchers widely agree that the ability to precisely retrieve desired information is vital to use the available knowledge. A way to achieve this is providing a retrieval system that is not only able to retrieve the available and sought information, but also to filter out irrelevant documents, while giving the relevant ones the highest ranking. The main goal of this work has been to investigate how to improve the ability for a system to find and rank relevant documents. Our method is based on applying series of information retrieval techniques to search in biomedical information and combine them in an optimal manner. These techniques include extending and using well-established information retrieval (IR) similarity models like the TF-IDF and BM25 as the scoring schemes, and applying personalisation so that researchers may affect the ranking based on their view of relevance. The techniques have been implemented and tested in a proof-of-concept prototype called BioTracer, extending a Java-based open source search engine library. The preliminary results from our experiments using the TREC 2004 Genomic Track collection seem satisfactory, with the best mean average precision (MAP) of 0.5129 and the best precision at 100 retrieved documents (P@100) of 0.473. What can be concluded from these results is that involving the users in the search will often have positive effects on the ranking of search results, and that our BioTracer system represents a tool that may be able to meet the user's information needs.