

Developing an IR and NLP-Based Algorithm for Biomedical Literature Searching

徐建業

Yen YT;Hsu CY

Abstract

MEDLINE is the largest biomedical journal database in the world. When researchers search for literatures by using Entrez system, they usually retrieve thousands of articles and the query results can only be sorted by PubMed ID, author, journal and publication date. However, users usually need to spend lots of time to review these articles to find out which they are really interested in. The purpose of this study is to apply Information Retrieval (IR) and Natural Language Processing (NLP) methods to develop an algorithm that can refine the search results of Entrez system. We have built a simple query system, the Biomedical Literature Searching System (BLSS) based on our algorithm. The system outperformed the Entrez for retrieving more relevant documents. In the future, we will build an application to help MEDLINE users retrieve articles that are most likely related to their queries.