Annotated Bibliography

Owan, V. J. (2022). A Data Mining Algorithm for Accessing Research Literature in Electronic Databases: Boolean Operators. In *Innovative Technologies for Enhancing Knowledge Access in Academic Libraries* (pp. 140-155). IGI Global.

This paper explores the importance of Boolean operators in the research to locate relevant materials. Research methods based on Boolean logic and Boolean operators were employed to locate a variety of medical materials. An example of the application of the Boolean operators is illustrated with a search strategy. The choice of a search strategy is dependent upon the research objective, the availability of the data, and the Boolean operators used. Boolean operators are used to specify complex Boolean queries over a collection of documents. A query written as a Boolean expression can be evaluated in the language of first-order logic using an effective algorithmic procedure. A search strategy, which is commonly used by information scientists to locate new information from literature, is proposed in this paper. It argues that Boolean logic is a suitable tool in search engines, but there are still few papers which use it. It gives an overview of techniques related to Boolean operators and their relevance to Internet databases. The technique of Boolean operators is explained and the importance of Boolean operators for indexing is discussed. It also describes the steps taken in the development of a Boolean-based search engine and gives examples.

Scells, H., Zuccon, G., & Koopman, B. (2019, May). Automatic Boolean query refinement for systematic review literature search. In *The world wide web conference* (pp. 1646-1656).

This paper explores the automatic Boolean query refinement for systematic review literature search. Query refinement helps improve precision but it also has potential to reduce recall, because the number of relevant documents decreases. A multi-perspective literature search was first conducted by using the Boolean search strategy. Subsequently, the output of the Boolean search result was subjected to refinement. Boolean query refinement is often implemented manually for the systematic review literature search. The existing Boolean query refinement approaches can be classified into four categories: syntax, pruning, statistics and data retrieval, and ranking. The syntactic query refinement can refine the query by removing unwanted words. The statistics query refinement can refine the query by including more relevant documents, such as documents containing query terms. The data retrieval query refinement can refine the query by including relevant documents based on citation information. The ranking query refinement can refine the query by ordering the retrieved documents. We discuss each of these query refinement approaches. In order to improve the automatic Boolean query refinement, we propose a framework, which includes two components, query evaluation component and query refinement component.

Vaidya, P., & Harinarayana, N. S. (2019). Comparison of User-generated Tags with Subject Descriptors, Author Keywords, and Title Terms of Scholarly Journal Articles: A Case Study of Marine Science. *Journal of Information Science Theory and Practice*, 7(1), 29-38.

This paper explores the comparison of user-generated tags with subject descriptors and author keywords. In order to assess the value of user-generated tags, we analyses their quality as well as their accuracy. Furthermore, we use these tags to create a semantic annotation dataset and to evaluate some of the existing approaches for the semantic annotation of content. Using subject descriptors and author keywords, we can evaluate the subject coverage, which represents the

extent of each paper's impact, the novelty of each paper and the amount of related topics of papers. The results of impact evaluation show that subject descriptors can provide a more comprehensive and accurate indication of the characteristics of papers than author keywords and the combined approaches. The results of the user study show that authors prefer using subject descriptors to author keywords, so we recommend using subject descriptors as the method to convey the content of a paper and provide a more accurate and complete evaluation for papers. The main result is that these descriptors, which traditionally have been treated as synonyms, are found to differ significantly in the number of tags they refer to. Therefore, they should not be treated as a single entity. Furthermore, subject descriptors have been found to be more predictive of the number of posts that will be found relevant, while author keywords are better indicators of the number of related posts.