Information Retrieval Systems

Abstract

An Information Retrieval (IR) system is an important and integral part of an information organization system; it involves the storing and retrieving of information which assist a user in conducting a query within a database. An IR system is the vehicle by which a user can locate records matching the terms of a query. This report enables readers to a) understand the purpose of an IR system, b) obtain a basic understanding of how such a system works, c) identify both system and human factors affecting the performance of such a system, and d) recognize how IR systems have improved the efficiency and accuracy of cataloging operations, benefiting both cataloger and user. Report examples demonstrate ways in which an IR system affects and improves cataloging of bibliographic information.
Introduction

This briefing enables the reader to: (1) understand the purpose of an information retrieval system, (2) obtain a basic understanding of how an information retrieval system works, (3) identify main factors which influence the performance of an information retrieval system, and (4) recognize how information retrieval systems have significantly improved the efficiency and accuracy of cataloging operations.

Definition

The information retrieval system is an essential component of an information organization system. It allows the user to search the representations of information objects as entered in a database. The IR system takes a user’s search terms, matches those terms with the information object representations, and displays (on a computer screen) the results.

Purpose and implications

Purpose of IR Systems

Information Retrieval systems allow a user to search for information objects which are part of an information organization system. The IR system stores metadata records and allows users various methods to search them. Retrieval methods vary between exact and partial match techniques. Most IR systems currently use exact match techniques, wherein the search terms have to exactly match the text information within the document in order to be retrieved. (Gloor, 1997).

How an IR System works

An IR system consists of three distinct files: full records; the dictionary, which alphabetically lists all unique words within the database; and the inversion list (Feldman, 1999). The first step is the entering of documents into the system, along with the building of an inverted file. New information objects are interfiled into the existing inversion list. Next, the query process takes place, in which a user types in some a term to be searched. Thirdly, the search terms are matched against the inverted file. Finally, the potential matching documents “are sorted by date, by field, or by how relevant the document is predicted to be to the query” (Feldman, 1999, Step 3: Query Matching section, ¶ 1). These records are then displayed for the user.

Factors Affecting Performance

Both system and human factors influence the performance of an IR system. System factors include: database, index language, indexing, and search strategy. The selection of a database is influenced by subject matter and cost factors. Index language includes the specificity of the vocabulary. Indexing refers to the level of exhaustivity within an information organization system. Search strategy is a factor associated with such concepts as Boolean operators, and natural language versus controlled vocabulary (Lancaster, 1993).

Human factors affecting IR system performance include: indexing, searching, and screening. Indexing is dependent on the level of subject expertise and experience of the cataloger responsible for entering the data into the database. Consistency among catalogers is another consideration in indexing. The searching methods of a system user also affect IR system performance. A user’s subject expertise, searching experience, and cognitive style determine search methods. Screening is the process by which a user decides what retrieved information is relevant to a search (Lancaster, 1993). It should be noted that “the systems do little to help the user understand and analyze what the system returns” (Feldman, 2000).

The Effect of IR Systems in Cataloging Operations
“...Information Retrieval Systems... have significantly improved the efficiency and accuracy of cataloging operations” (Jian, 1997). Electronic databases allow catalogers to input data into information organization systems much more quickly and efficiently than was possible with hand-written or typed entries on cards, as used in non-electronic card catalog systems. The input of bibliographic information is made more consistent and accurate with the development of IR systems.

Examples

Example #1: Indexing of terms within an IR system provide consistency.

Both catalogers and end users are benefited by the creation of a mechanism which indexes terms in a database. A cataloger may not be familiar with subject terminology, for instance, and has a less difficult job of entering subject terms due to the thesaurus within an IR system. Furthermore, a user conducting a query is not required to input every term which could be related to a search. For example, a user types in the term car, and the IR system pulls up records containing the subject terms car, automobile, and vehicle.

Example #2: Accuracy of records is improved with an IR system.

A cataloger is able to get accurate records from internet databases to aid in the creation of records being entered into an organization system. The indexer can access records, using information retrieval, from other databases, including online sources such as Library of Congress. Thus, information such as variations of an author's name for use in a database's name authority file, is attainable with IR systems.

Example #3: IR systems allow for differences in searching strategies of users.

Formerly, when cataloging involved hand-written or typed cards which were placed in a set of file drawers, users could search only by subject, title, and author. An information retrieval system allows users to search by a wide variety of additional methods, including: keyword, publisher, date of publication, genre, series, and ISBN. Additionally, IR systems allow users the ability to browse, as defined by Taylor (2004) as “scan lists of terms, headings, or brief titles to find topics...of interest” (p. 113), quicker and more efficiently than was possible within the confines of a traditional paper card catalog.

Conclusion

The information retrieval system within an information organization system allows a user to efficiently search a database for records by matching search terms with data about information objects as entered into the IR system. Three distinct files—full records, dictionary, and inversion list—make up an IR system. Search terms are matched with data representing bibliographic records, and displayed for the user. A number of both system and human factors contribute to the performance of an IR system. These factors include: selection of database, index language, indexing, and search strategy; as well as indexing, searching, and screening capabilities on the part of the user. IR systems have significantly improved efficiency and accuracy in the realm of library and information science.
References


