

INFORMATION RETRIEVAL AND THE ASSOCIATED PROBLEMS

S. Raman Venkataraman
Science Division, Library
University of British Columbia
Vancouver, B.C.

ABSTRACTS

Two decades have passed since the introduction of computer-based information retrieval systems. Online information retrieval systems offer searchers enormous advantages over traditional methods. Simultaneously, it has introduced new issues and problems. Many attribute these problems to information system failure. One of the measures of failure is Expectation Failure, i.e. the inability of an information retrieval (IR) system to meet user expectations. It is measured in terms of quality and comprehensiveness of the product retrieved by the system. The product is in turn dependent on the quality of the database. The quality of the database is evaluated on the basis of the following criteria: 1. Coverage, 2. Errors in entry, 3. Duplication or overlap, 4. Availability of controlled vocabulary and 5. Standardization of retrieval techniques.

In this paper I shall first discuss the shortcomings of databases, such as (a) coverage (b) duplication or overlap (c) lack of comprehensiveness (d) spelling inconsistency and (e) errors in database records. Secondly, I shall review various measures undertaken by database producers and online hosts to solve some of these shortcomings. Finally, I shall discuss CD-ROM databases, which have provided a new dimension to information retrieval systems.

Locating reliable data about indexing and abstracting coverage of sources in a database is often difficult. It is evident from the published data on coverage and online search results that databases vary in scope and selection of articles from the sources scanned. Many databases are selective and uneven in their coverage. Although, it is indicated that a database contains abstracts from a specific number of journals cover to cover, a random test showed that this condition is not always fulfilled.

Due to uneven subject coverage by databases, there is no single database which can be relied upon as the comprehensive source for many subjects, especially interdisciplinary subjects. So there is a need to search more than one database for a comprehensive or exhaustive search on a topic.

Overlap will be found among retrievals from different databases, provided the databases cover the same sources and include in the database the same record or records from the same source.

Due to lack of standardization, for a comprehensive search one should use both American and English spellings. Some records in a database may have both American and English spellings for the same word, which may result in retrieving duplicate records.

Computer searching of natural language databases require searchers to be aware of spelling errors in many databases.

Many online vendors have introduced, besides BOOLEAN operators, the following search features: Proximity, Truncation, and "Expand". But the online services differ in their search features.

CD-ROM databases have provided a new dimension to the information retrieval system. One of the advantages in searching CD-ROM databases is that it has no connect time charge. However, it has its own limitations.