Chapter 5

Information Analysis in Digital Library Environments: Lessons Learned in Pharma

Lauren Harrison
Roche Pharmaceutical Corp., USA

ABSTRACT

This chapter addresses the question of how the analysis of results retrieved from online bibliographic information systems changed over the last 32 years as digital libraries have evolved. It demonstrates that Digital Libraries of the future will enable knowledge discovery by providing direct access to the semantic content of documents through the implementation of text mining tools. To achieve this research with IR systems and text-mining tools, pipeline pilot (Bandy, et al., 2009), I2E (Vellay, 2009), and BioText will need to be conducted by experts in information retrieval not just subject scientific specialists.

INTRODUCTION

The process of searching and retrieving relevant data from bibliographic databases has relatively stayed the same. In the beginning Boolean logic and proximity searching (right and latter on left hand) are applied to online systems containing records that were created as surrogates for a publication/article. Indexing and thesauri were created to facilitate the identification of relevant articles. This has not changed. Experimentation with full text bibliographic databases resulted in statistical algorithms to rank search results. None of these relevance-ranking mechanisms were not adopted by the major vendors—Dialog, DataStar, STN.

Today, most online bibliographic database collections are now linked to the institutions licensed e-journal collection so that the database

DOI: 10.4018/978-1-4666-2991-2.ch005
record is directly linked to the pdf or html of the document it represents. This affords opportunities for experimentation in the post processing of information results not previously available. This is the greatest change in information analysis.

HISTORY OF POST PROCESSING OF RETRIEVAL RESULTS

In the beginning there was manual “cut and paste.” The database records contained the text of a bibliographic citation—the title, author, journal, and keywords of the referenced article. A Boolean query matched specific words to the words in the fields of the database record. The hardware capacity of the mainframes of this period determined the functionality of the IR system. The corpus was a collection of database records rather than a collection of the full text articles. The output device was a paper teletypewriter. The network was a telephone line with a coupler. These systems were best suited to generating a bibliography—an exhaustive printout of all articles that contained the key words on a subject such as diabetes and insulin. These collections of records that online systems handled became known as bibliographic databases. These bibliographies databases (e.g. Medline in Biology and Medicine and STN in Chemistry) grew over time to include searchable abstracts of the articles as a result of the growth of the indexing and abstracting industry.

Harrison and Lacerna (1992) documented the process of producing annotated bibliographies in Pharma prior to 1989. A considerable amount of production time was spent in the content analysis and review of the abstracts or original articles to determine various topics relevant to the drug profile. Prior to 1989, the process by which bibliographies were created required manual manipulation of the exported in-house database records. A print out of all records on a substance was retrieved. Each record was then manually coded or indexed according to the categories.

If a record was indexed by two or more categories, two or more photocopies were made. References were then cut, pasted and sorted manually into alphabetical lists per category. The resulting report listed records alphabetically by author within each category. This process had to be repeated per category and each category was input in the logical order of appearance desired. An author index indicating the location of authors works within the subject section could be automatically generated. This process was tedious, laborious, and costly in terms of man-hours spent in compilation.

In 1987 research on an automated method of bibliography general using Sci-Mate Manager was initiated. This initiative went on to review all bibliographic manipulation software available at the time. Pro-Cite was ultimate chosen as the best tool to facilitate bibliography generation as on a daily basis it eliminated the need to edit downloaded data for search report generation. It provided an efficient means of electronically combining references from several sources, facilitating the review of data by providing duplicate detection assistance. It greatly reduced the amount of post processing required to produce the subject section of the bibliography. Most importantly, at the time, the application was programmer independent and the users (information scientists) had total control over the output appearance.

THE ADVENT OF PERSONAL VIRTUAL LIBRARIES

Since then many other tools of this genre have appeared—Endnotes, Ref works, References Manager to name a few. All have attributes comparable to Pro-Cite. However, the most recent tool in this arena, Quosa Platinum Manager, has a decided advantage. It was designed in the digital library age and it not only facilitates the import/download of bibliographic database records but it retrieves the pdf or html format of the full text
4 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage:
www.igi-global.com/chapter/information-analysis-digital-library-environments/74323?camid=4v1

www.igi-global.com/e-resources/library-recommendation/?id=1

Related Content

A Probabilistic SVM Approach to Annotation of Calcification Mammograms
www.igi-global.com/article/probabilistic-svm-approach-annotation-calcification/45734?camid=4v1a

Decomposed PRNU Library for Forensics on Photos
www.igi-global.com/article/decomposed-prnu-library-forensics-photos/51651?camid=4v1a

Managing Change in Reference and Information Services in Digital Environment
www.igi-global.com/chapter/managing-change-reference-information-services/42742?camid=4v1a

Design and Development of Institutional Repository Using Dspace: A Case Study of Kuvempu University, India
www.igi-global.com/article/design-and-development-of-institutional-repository-using-dspace/83498?camid=4v1a