

Information Retrieval Library (IRLIB)

Carolyn Schmidt

National Institute of Standards & Technology
100 Bureau Drive, Stop 8940
Gaithersburg, MD 20899-8940
01-301-975-3243

carolyn.schmidt@nist.gov

Donna Harman

National Institute of Standards & Technology
100 Bureau Drive, Stop 8940
Gaithersburg, MD 20899-8940
01-301-975-3569

donna.harman@nist.gov

ABSTRACT

The IRLIB is a full-text web searchable digital library, whose contents are publications and documents relative to the field of information retrieval.

Keywords

Digital library, archival documents

1. DESCRIPTION

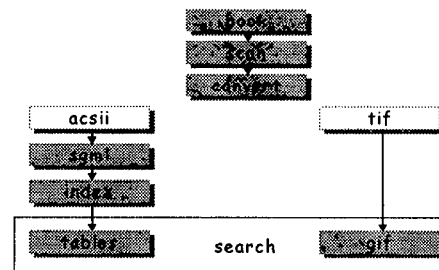
The IRLIB goals are three-fold: (1) to preserve access to non-electronic information retrieval related research resources; (2) to research digital library technologies and user interface issues; and (3) to eventually collect queries on-line from Information Retrieval (IR) and Digital Library (DL) literate communities for further use in research.

The IRLIB currently contains six government-funded publications and one privately published book, as follows:

1. ISR-10 (March 1996), the full PhD Thesis of Joseph John Rocchio, Jr.
2. ISR-11 (June 1966), containing early reports from the SMART project
3. IRS-13 (December 1967), containing reports on new evaluations, in particular work by Michael Keen
4. NIST Monograph 91 (March 1965), containing a survey of the literature on automatic indexing (182 pages plus an extensive bibliography)
5. The Second Text Retrieval Conference (TREC-2)
6. The First Text Retrieval Conference (TREC-1)
7. Information Retrieval Experiment (1981), book by Karen Sparck Jones

The library was created by defining several processes that convert a physical document to electronic form. The search index was created by NIST's ZPRISE system, and retrieval of output is done using customized CGI scripts. The document pages are displayed both as gif images and as appended OCR output. The current interface is a traditional representation of ranked lists in response to natural language queries, but other interfaces will be developed in the future to make more use of the metadata within the publications.

Processes of Building the IRLIB



2. CONCLUSION

NIST is presenting this system both to alert the community as to this new resource and to get suggestions for further work. One obvious place for further work is the interface, but it is not clear what changes are needed for searching full-text documents where the full text includes entire books that are available as OCR images only. There is the additional issue of how people, in particular the scientific community involved in IR, would prefer to access this material in addition to traditional word-search mechanisms.

The IRLIB is accessible via an internet browser at <http://www.nist.gov/itl/div894/894.02/projects/irlib>