

Running head: SUBJECT HEADINGS FOR A DATABASE

Subject Headings for a Database:

Developing the Controlled Vocabulary

Kristin Yiotis, Zorica Kovacevic, and Chad Harris

San Jose State University

LIBR 202.01

Information Retrieval

Professor: Judy Weedman

May 24, 2004

APA Style

Subject Headings for a Database:
Developing the Controlled Vocabulary

PART A

User Guide

Information about the database

Scope of the database

At this time the database consists of a collection of 15 documents relating to information storage and retrieval. The documents, chosen at random, present a broad range of possible research topics for the user group, first year library and information science students. This database allows students to search by subject and by descriptive data for articles related to topics they have chosen for final papers in LIBR 200, an introductory library science class.

The user group has a basic level of experience with the subject area and retrieval systems. This is the first or near first class these students have taken in their program. Many students have worked in libraries but they might have only basic knowledge of the research side of library science. The students know the subject area of information retrieval in a practical, hands on, sense but not in a theoretical or conceptual sense. Their knowledge of retrieval systems is also practical, hands on, knowledge based on using library online catalogs and commercial and library databases. The users have basic understanding of InMagic DBTextWorks.

The currency of the database is from the years 1981 to 1998, the dates the documents were published as journal articles, with the majority coming from the early 1990s. Many are not available in online library databases and are only available in this

database. All but one of these articles were published in scholarly journals in the field of library and information science, the majority being library and information science professional association publications, such as the Journal of the American Society for Information Science and Technology. One of the documents is a chapter from a book related to information science. This database is updated on a yearly basis in January when the prior year's articles in the represented journals become available.

Good introduction to and description of the database.

Guidelines that enable searching

What users need to know to search the database effectively ? [As it stands, this is not a question – you either need to add the word “do” after “What” to make it a question, or remove the question mark.](#)

This database consists of fields. Some fields are descriptive information that identifies a specific article. Users search these descriptive fields using "natural" grammatical English or numbers. A user would search by descriptive data if she/he already knew the title, author, publication, year, citation, or document number of the article she/he wanted to retrieve. Users can also search by subject. A user would search by subject if she/he needed to locate articles about a particular subject. The fields that enable searching by subjects are the title, abstract, and vocabulary fields. While the title and abstract fields exactly describe specific articles, they are also useful for searching subjects because an article's topic is usually embodied in its title and abstracts are created specifically to summarize an article's main concepts. [good; this is very clear](#)

The vocabulary fields are designed to enable a user to search by subject using a list of authorized terms (subject headings) to predictably find documents that are about a particular subject. Subject headings indicate major concepts or subjects found in each

article. This database uses two kinds of subject headings to represent the concepts of the articles it collects. One is a precoordinate vocabulary and the other is a postcoordinate vocabulary, and each has a separate validation list for its appropriate field. The precoordinate vocabulary consists of major headings and subheadings, like topics and subtopics. Some examples are digital documents--standards, information retrieval--searching, online information systems--theories. The postcoordinate vocabulary consists of single terms only, with no distinction between topic or subtopic. Some examples are digital documents, information retrieval, online information systems, searching, standards, and theories. All fields allow for word and term (phrase) searches except the numbers fields, doc number and year, which permit only numbers.

The controlled vocabulary fields are term indexed, which requires that entire term is entered as the value and not a part of the term (one other possibility is that only the beginning part of the term can be entered, and truncated). The controlled vocabulary is designed in inverted word order, not natural language order, a decision made by the designers based on the concepts of user warrant vs. literary warrant.

What users need to know about the vocabulary fields in general ?

The natural language fields replicate the documents exactly, such that title, author, publication, and abstract match that of the article. The controlled vocabulary fields are derived fields, such that the subject headings represent the concepts found in the articles, and a user can only search for the terms that are in control vocabularies. Natural language fields in title and abstract are good places to start searching if the user wants to immediately get into the records. These fields serve as gateways to the contents of the database using terms familiar to a user, and a user has to think of the possibilities because

natural language is not controlled for what she/he can search. The user can enter keywords in the natural language fields to retrieve documents with these words in the title or abstract. Keywords are any words **in other** than the stop words that are listed below.

The descriptive data fields are most useful when the user wants to retrieve a specific article or set of articles by title, author, publication, or year. The title field is also useful for subject searching. The database uses the Boolean operators, AND, NOT, OR, between fields for combined field searches and combines word or phrase searches within fields using symbols instead: & for AND, / for OR and ! for NOT. The database allows for searching ranges of dates using the colon between years, e.g. 1980:1988, and dates up to, after, and including, using <, >, =, such as <=1990, up to and including 1990. Multiple years can be searched in a single query using Boolean OR only /, which will pull up all years. The year field does not truncate.

When searching the controlled vocabulary fields a user needs to go to the validation lists and look-up the terms for the concepts she/he is looking for. In the query window, a user can browse the controlled vocabulary lists by entering the cursor in the pre- or post-co vocabulary fields and using the F3 key to find the validation lists. If the user enters a term that is not one of the controlled vocabularies and gets no results, then she/he explores the entry vocabulary for the term used in the controlled vocabulary. All natural language and controlled vocabulary fields allow word and term truncation to expand the search to all forms of the root word. For example index* includes indexing, indexes, index, and indexers. Inform* includes informal, informational, etc. The controlled vocabulary fields allow only term searches in order to enforce a distinction between the precoordinate and postcoordinate vocabularies. If they allowed word

searches the two fields would be virtually identical and the search in pre-co would be returned to post-co field. Also, the precision and specificity that precoordinate vocabulary gives would be lost.

A user may begin searching concepts as keyword searches a natural language field to aggregate all documents on a subject and then combine fields to refine the search, moving from recall to precision by adding another variable. For instance the user may want articles about information retrieval that were published before and including 1985 to explore early information retrieval systems or after and including 2000 to explore recent developments.

The database uses Boolean operators when combining fields in the database. AND restricts the search by requiring that values in both fields be included in the documents retrieved achieving greater precision. OR expands the search to by allowing either value to be in the documents retrieved achieving greater recall. A combined search using keywords in a natural language field plus a subject heading in a controlled vocabulary field using the Boolean OR would expand the search and increase recall, while the Boolean AND would limit the search and increase precision. Boolean NOT would also increase precision.

This section is very thorough and very useful.

Rules that govern the database

1. Standards. Organizing principles govern the syntax of the vocabulary in this database. The syntax is organized hierarchically, moving from the general to the specific. Thus the broader term precedes the narrower term in an inverted word order. [Is this true of only the postco, or postco and preco both? \(I can look, but it would be useful to say.\)](#)

I didn't actually see any inverted word order in your validation lists.

2. Capitalization and punctuation. The database ignores all punctuation and capitalization. The field values follow standard English grammar, such that only proper nouns are capitalized.
3. Grammar. When subject field vocabularies are nouns the database uses the plural form, except for mass nouns. When verbs the database uses the –ing form, such as indexing. Truncating at the root word eliminates the need to predict grammatical form.
4. Names of people. People’s names have been included in the subject fields because such people are the subjects of articles. Users can search in the controlled vocabulary fields for people by name in this fashion: Last, First, M., Last, F., e.g. Boole, George.
5. Stop words. Natural language fields will search for any keyword except stop words, such as:

a, an, and, by, for, from, in, of, the, to.

6. Entry vocabulary. This list of entry vocabulary anticipates that the database users are introductory library science students whose vocabulary might differ from the vocabulary derived from the articles (user warrant vs. literary warrant). The rationale for an entry vocabulary is that users will have varying levels of expertise in the subject area and will need redirecting to the authorized vocabulary, sort of an entry to the control vocabulary, a starting point for the subject they are looking for. **Very useful!**

Entry vocabulary, precoordinate

- Cognitive science use information processing—human cognition
- Computer databases use online information systems--
- Controlled vocabulary use search terms—descriptors, subject assess—thesauri
- Interview techniques use reference interviews—questioning strategies
- Online catalogs use online information systems--
- Systems design use online information systems--designs

Entry vocabulary, postcoordinate

This would be easier to read/understand if you did something with the font to distinguish the entry terms from the authorized terms, or put the “use” in a different font.

- Cognitive science use human cognition
- Computer databases use online information systems
- Controlled vocabulary use descriptors, search terms, subject assess, thesauri I assume you mean that any of these might be an appropriate substitute for *controlled vocabulary*, depending on the particular interests of the searcher.
- Interview techniques use questioning strategies, reference interviews
- Online catalogs use online information systems This is good for aggregation but bad for discrimination; online catalogs are one very specific kind of online information system, and not using it as a subject means that users can't search just for online catalogs without getting other systems like the Internet, databases, etc., as well.
- Systems design use designs, online information systems I looked and looked at this and finally decided designs is one term and online information systems is another. Does this mean that the searcher should phrase the query as *designs & online information systems*? If so, it would be useful to show the Boolean logic rather than a comma. When I looked at it, I thought it was a single term with the natural word order inverted: *designs, online information systems*

7. The authorized vocabulary fields. They are term indexed, which requires that entire term is entered as the value and not a part of the term. The controlled vocabulary is designed in inverted word order, not natural language order, a decision made by the designers based on a consideration of what was most useful to the user (user warrant) and the logic of the particular concepts in the article (literary warrant).

The Validation List for Precoordinate and Postcoordinate Fields

Term index for field 'PRECO', textbase 'SubjectAnalysisDB', 5/23/2004 12:49:37 PM:

- 1 Boole George
- 1 digital documents preservation
- 1 digital documents standards
- 2 human computer interaction
- 2 information processing human cognition
- 1 information retrieval Boolean algebra
- 1 Information Retrieval Document Ranking
- 3 Information Retrieval Document Relevance
- 1 Information Retrieval Document Selection
- 1 information retrieval human factors
- 2 Information Retrieval Indexing
- 1 information retrieval models
- 7 Information Retrieval Searching
- 2 information retrieval theories

Capitalization doesn't matter for searching, but your vocabulary list looks better if you're consistent.

- 1 information retrieval user studies
- 3 information seeking behavior
- 1 online information systems access points
- 1 online information systems artificial intelligence
- 5 online information systems design You say designs in your entry vocab.
- 2 online information systems evaluation
- 1 online information systems history
- 1 online information systems human factors
- 1 online information systems interfaces
- 2 online information systems needs assessment
- 1 online information systems statistics
- 6 online information systems user studies
- 1 psychology memory
- 3 psychology theories
- 1 reference interviews
- 1 reference interviews questioning strategies
- 1 search terms descriptors
- 1 search terms subject headings
- 1 Shannon Claude
- 2 subject access thesauri
- 1 Taube Mortimer

Good aggregation.

Total number of keys: 35

Term index for field 'POSTCO', textbase 'SubjectAnalysisDB', 5/23/2004 12:43:40 PM:

- 1 access points
- 1 artificial intelligence
- 1 Boole George
- 1 Boolean algebra
- 1 descriptors
- 5 design
- 1 digital documents
- 1 document ranking
- 3 document relevance
- 1 document selection
- 2 evaluation
- 1 history
- 2 human cognition
- 2 human computer interaction
- 2 human factors
- 2 indexing
- 2 information processing
- 11 information retrieval
- 3 information seeking behavior
- 1 interfaces
- 1 memory

- 1 models
- 2 needs assessment
- 10 online information systems
- 1 preservation
- 3 psychology
- 1 questioning strategies
- 1 reference interviews
- 2 search terms
- 7 searching
- 1 Shannon Claude
- 1 standards
- 1 statistics
- 2 subject access
- 1 subject headings
- 1 Taube Mortimer
- 4 theories
- 2 thesauri
- 6 user studies

Total number of keys: 39

The Data Structure

Textbase: SubjectAnalysisDB
Created: 4/10/2004 6:19:43 PM
Modified: 4/17/2004 8:29:29 PM

Field Summary:

- 1. DOC_NO: Automatic Number(next avail=16, increm=1), Term
- 2. AUTHOR: Text, Term & Word
- 3. TITLE: Text, Term & Word
- 4. CITE: Text, Term & Word
- 5. ABSTRACT: Text, Word
- 6. POSTCO: Text, Term
- 7. PRECO: Text, Term
- 8. PUBLICATION: Text, Term & Word
- 9. YEAR: Number, Term

Both vocabularies are well structured and provide good aggregation. (I don't see any of the inverted word order you referred to earlier.) Later comment: after reading your discussion section, I understood better what you meant by "inverted word order." But I think the real issue may be which concept is most fundamental rather than the verb-direct object relationship.

good

Log file enabled, showing 'DOC_NO'

Leading articles: a an the

Stop words: a an and by for from in of the to

Textbase Defaults:

Default indexing mode: SHARED IMMEDIATE

Default sort order: <none>

Textbase passwords:

Master password = "
 0 Access passwords:
 No Silent password

The Records

DOC_NO 1

AUTHOR Bates, Marcia J.

TITLE Subject Access in Online Catalogs: A Design Model

CITE v. 37, no. 6 (November), 357-375

ABSTRACT A model based on strikingly different philosophical assumptions from those currently popular is proposed for the design of online subject catalog access. Three design principles are presented and discussed: uncertainty (subject indexing is indeterminate and probabilistic beyond a certain point), variety (by Ashby's law of requisite variety, variety of searcher query must equal variety of document indexing), and complexity (the search process, particularly during the entry and orientation phases, is subtler and more complex, on several grounds, than current models assume). Design features presented are an access phase, including entry and orientation, a hunting phase, and a selection phase. An end-user thesaurus and a front-end system mind are presented as examples of online catalog system components to improve searcher success during entry and orientation.

The proposed model is "wrapped around" existing Library of Congress subject-heading indexing in such a way as to enhance access greatly without requiring reindexing. It is argued that both for cost reasons and in principle this is a superior approach to other design philosophies.

POSTCO design
 indexing
 information retrieval
 online information systems
 models
 search terms
 searching
 subject access
 subject headings
 thesauri
 user studies

PRECO information retrieval--indexing
 information retrieval--models
 information retrieval--searching
 online information systems--user studies
 online information systems--design
 search terms--subject headings
 subject access--thesauri

PUBLICATION Journal of the American Society for Information Science

YEAR 1986

DOC_NO 2

AUTHOR Dervin, Brenda & Dewdney, Patricia

TITLE Neutral Questioning: A new approach to the reference interview

CITE Summer, 506-513

ABSTRACT Neutral questioning is a strategy for conducting the reference interview in a way that allows the librarian to understand

the query from the user's viewpoint. Neutral questions are open in form, avoid premature diagnosis of the problem, and structure the interview along dimensions important to the users. Derived from extensive studies into information-seeking behavior, this strategy has now been taught to over a thousand practitioners through workshops developed in 1981. Empirically based research to test the effects of neutral questioning is in progress. Informal evaluation indicates that neutral questioning may become a useful component of in-service training for librarians.

POSTCO human factors
information retrieval
questioning strategies
reference interviews

PRECO information retrieval--human factors
reference interviews
reference interviews--questioning strategies

PUBLICATION RQ

YEAR 1986

DOC_NO 3

AUTHOR Rothenberg, Jeff

TITLE Ensuring the Longevity of Digital Documents

CITE v. 272, no. 1 (January), 42-47

ABSTRACT The digital medium is replacing paper in a dramatic record-keeping revolution. But such documents may be lost unless we act now. Information technology is revolutionizing our concept of record keeping in an upheaval as great as the introduction of printing, if not of writing itself. The current generation of digital records has unique historical significance. Yet these documents are far more fragile than paper, placing the chronicle of our entire period in jeopardy. The significance of many digital documents--those we consider too unimportant to archive--may become apparent only long after they become unreadable. Many of the traditional methods developed for archiving printed matter are not applicable to electronic files. The content and historical value of thousands of records, databases, and personal documents may be irretrievably lost to future generations if we do not take steps to preserve them now.

POSTCO digital documents
preservation
standards

PRECO digital documents--preservation
digital documents--standards

PUBLICATION Scientific American

YEAR 1995

DOC_NO 4

AUTHOR Simon, Herbert A.

TITLE Information-Processing Models of Cognition

CITE September, no. 5, 364-377

ABSTRACT This article reviews recent progress in modeling human cognitive processes. Particular attention is paid to the use of computer programming languages as a formalism for modeling, and to computer simulation of the behavior of the systems modeled. Theories of human cognitive processes can be attempted at several levels: at the level of neural processes, at the level of elementary information processes (e.g.,

retrieval from memory, scanning down lists in memory, comparing simple symbols, etc.), or at the level of higher mental processes (e.g., problem solving, concept attainment). This article will not deal at all with neural models; it focuses mainly upon higher mental processes, but not without some attention to modeling the elementary processes and especially to the relationships between elementary and complex processes.

POSTCO human cognition
information processing
psychology
theories

PRECO information processing--human cognition
psychology--theories

PUBLICATION Journal of the American Society for Information Science

YEAR 1981

DOC_NO 5

AUTHOR Harter, Stephen & Cheng, Yung-Rang

TITLE Colinked Descriptors: Improving Vocabulary Selection for End-User Searching

CITE v. 47, no. 4 (April), 311-325

ABSTRACT This article introduces a new concept and technique for information retrieval called colinked descriptors. Borrowed from an analogous idea in bibliometrics--cited references--colinked descriptors provide a theory and method for identifying search terms that, by hypothesis, will be superior to those entered initially by a searcher. The theory suggests a means of moving automatically from two or more initial search terms, to other terms that should be superior in retrieval performance to the two original terms. A research project designed to test this colinked descriptor hypothesis is reported. The results suggest that the approach is effective, although methodological problems in testing the idea are reported. Algorithms to generate colinked descriptors can be incorporated easily into system interfaces, front-end or pre-search systems, or help software, in any database that employs a thesaurus. The potential use of colinked descriptors is a strong argument for building richer and more complex thesauri that reflect as many legitimate links among descriptors as possible.

POSTCO descriptors
document relevance
information retrieval
online information systems
search terms
subject access
thesauri
user studies

PRECO information retrieval--document relevance
online information systems--user studies
search terms--descriptors
subject access--thesauri

PUBLICATION Journal of the American Society for Information Science

YEAR 1996

DOC_NO 6

AUTHOR Harter, Stephen P.

TITLE Psychological relevance and information science

CITE v. 43, no. 9 (October) 602-615

ABSTRACT This article summarizes the theory of psychological relevance proposed by Dan Sperber and Deirdre Wilson (1986), to explicate the relevance of speech utterances to hearers in everyday conversation. The theory is then interpreted as the concept of relevance in information retrieval, and an extended example is presented. Implications of psychological relevance for research in information retrieval; evaluation of information retrieval systems; and the concepts of information, information need, and the information-seeking process are explored. Connections of the theory to ideas in bibliometrics are also suggested.

POSTCO document relevance
information seeking behavior
information retrieval
psychology
searching
theories

PRECO information retrieval--document relevance
information retrieval--searching
information retrieval--theories
information seeking behavior
psychology--theories

PUBLICATION Journal of the American Society for Information Science
YEAR 1992

DOC_NO 7

AUTHOR Gauch, Susan

TITLE Intelligent Information Retrieval: An Introduction

CITE v. 43, no. 2 (March), 175-182

ABSTRACT Researchers are exploring the application of artificial intelligence techniques to information retrieval with the goal of providing intelligent access to online information. This article surveys several such systems to show what is possible in the lab today, and what may be possible in the library or office of tomorrow. Systems incorporating user modeling, natural language understanding, and expert systems technology are presented.

POSTCO artificial intelligence
human-computer interaction
information retrieval
interfaces
online information systems
searching

PRECO human-computer interaction
information retrieval--searching
online information systems--artificial intelligence
online information systems--interfaces

PUBLICATION Journal of the American Society for Information Science
YEAR 1992

DOC_NO 8

AUTHOR Huston, Mary M.

TITLE Windows into the Search Process: An Inquiry into Dimensions of Online Information Retrieval

CITE v. 19, no. 3/4 (June/August), 227-243

ABSTRACT From diverse users' points of view, contextual frameworks are elaborated for the nature of the information technology, the information universe, and the information search. Within these conceptual

parameters, established theories on search strategy are reviewed and cognitive models of information-seeking are highlighted. Future directions for research on users' search processes are discussed in terms of the role for online retrieval in the future information environment.

POSTCO access points
 information retrieval
 information seeking behavior
 needs assessment
 online information systems
 searching
 user studies

PRECO information seeking behavior
 information retrieval--searching
 online information systems--access points
 online information systems--needs assessment
 online information systems--user studies

PUBLICATION Online Review

YEAR 1991

DOC_NO 9

AUTHOR Najarian, Suzanne E.

TITLE Organizational Factors in Human Memory: Implications for Library Organization and Access Systems

CITE v. 51, no. 3, 269-291

ABSTRACT Psychological studies on memory and learning are examined for what they might reveal about human categorizing processes and the organizing principles and limitations of human memory. Particular attention is given to evidence for a model which represents the organization of knowledge in memory in terms of a hierarchical type of structure. The experimental findings suggest several considerations for the design of library systems of organization and access that would take into account characteristics of the conceptual organization of knowledge. Such systems are likely to be particularly effective in aiding the user in his search for information since they would (1) employ organizational schemes that are familiar to the individual, (2) permit a strategy for the exploration of a subject area similar to the type of search procedure which seems to facilitate the retrieval of items from memory, and (3) take into consideration the apparent limits on the amount of information that the individual can successfully attend to at one time.

POSTCO design
 human cognition
 information processing
 memory
 online information systems
 psychology
 theories

PRECO information processing--human cognition
 online information systems--design
 psychology--memory
 psychology--theories

PUBLICATION The Library Quarterly

YEAR 1981

DOC_NO 10

AUTHOR Soergel, Dagobert

TITLE Assessment of User's Problems and Needs

CITE 93-107 (Chapter 7)

ABSTRACT The importance of studying user needs as a basis for information system evaluation and design is discussed, and several approaches to the study of user needs are discussed, including the use of user feedback, examination of past use, and problem analysis based on records.

POSTCO design
evaluation
needs assessment
online information systems
user studies

PRECO information retrieval--user studies
online information systems--design
online information systems--evaluation
online information systems--needs assessment
online information systems--user studies

PUBLICATION Organizing Information: Principles of Database and Retrieval Systems

YEAR 1985

DOC_NO 11

AUTHOR Bates, Marcia J.

TITLE Indexing and Access for Digital Libraries and the Internet: Human, Database, and Domain Factors

CITE v. 49, no. 13, 1185-1205

ABSTRACT Discussion in the research community and among the general public regarding content indexing (especially subject indexing) and access to digital resources, especially on the Internet, has underutilized research on a variety of factors that are important in the design of such access mechanisms. Some of these factors and issues are reviewed and implications drawn for information system design in the era of electronic access. Specifically the following are discussed: Human factors: Subject searching vs. indexing, multiple terms of access, folk classification, basic-level terms, and folk access; Database factors: Bradford's Law, vocabulary scalability, the Resnikoff-Dolby 30:1 Rule; Domain factors: Role of domain in indexing.

POSTCO design
human factors
information retrieval
indexing
online information systems
searching

PRECO information retrieval--indexing
information retrieval--searching
online information systems--design
online information systems--human factors

PUBLICATION Journal of the American Society for Information Science

YEAR 1998

DOC_NO 12

AUTHOR Maron, M. E.

TITLE Probabilistic Design Principles for Conventional and Full-Text

Retrieval Systems

CITE v. 24, no. 3, 249-255

ABSTRACT In order for conventionally designed commercial document retrieval systems to perform perfectly, the following two (logical) conditions must be satisfied for every search: (1) There exists a document property (or combination of properties) that belongs to those (and only those) documents that are relevant. (2) That property (or combination of properties) can be correctly guessed by the searcher. In general, the first assumption is false, and the second is impossible to satisfy; hence no conventional IR system can perform at a maximum level of effectiveness. (We are painfully aware of the current poor performance values for Recall and Precision. Furthermore, Recall deteriorates rapidly as document corpora continue to grow in size.) However, different design principles can lead to improved performance. This article presents a view of the document retrieval problem that shows that since the relationship between document properties (whether they be humanly assigned index terms or words that occur in the running text) and relevance is at best probabilistic, one should approach the design problem using probabilistic principles. It turns out that a front end designed to permit searchers to attach probabilistically interpreted weights to their query terms could be adapted for conventional IR systems. Such an enhancement could lead to improved performance.

POSTCO design
document relevance
document ranking
evaluation
information retrieval
searching
statistics

PRECO information retrieval--document ranking
information retrieval--document relevance
information retrieval--searching
online information systems--design
online information systems--evaluation
online information systems--statistics

PUBLICATION Information Processing and Management**YEAR** 1988**DOC_NO** 13**AUTHOR** Smith, Elizabeth S.

TITLE On the Shoulders of Giants: From Boole to Shannon to Taube: The Origins and Development of Computerized Information from the Mid-19th Century to the Present

CITE v. 12, no.2 (June), 217-226

ABSTRACT This article describes the evolution of computerized information storage and retrieval, from its beginnings in the theoretical works on logic by George Boole in the mid-nineteenth century, to the application of Boole's logic to switching circuits by Claude Shannon in the late 1930s, and the development of coordinate indexing by Mortimer Taube in the late 1940s and early 1950s. Thus, electronic storage and retrieval of information, as we know it today, was the result of two major achievements: the advancement of computer technology initiated to a large extent by the work of Shannon, and the development of coordinate indexing and retrieval by the work of Taube. Both these achievements are based on and are the application of the theoretical works of George Boole.

POSTCO Boole, George
 Boolean algebra
 history
 information retrieval
 online information systems
 Shannon, Claude
 Taube, Mortimer
 theories

PRECO Boole, George
 information retrieval--Boolean algebra
 information retrieval--theories
 online information systems--history
 Shannon, Claude
 Taube, Mortimer

PUBLICATION Information Technology and Libraries
YEAR 1993

DOC_NO 14
AUTHOR Spink, Amanda
TITLE Term Relevance Feedback and Mediated Database Searching:
 Implications for Information Retrieval Practice and System Design
CITE v. 31, no. 2, 161-171
ABSTRACT Research into both the algorithmic and human approaches to information retrieval is required to improve information retrieval system design and database searching effectiveness. This study uses the human approach to examine the sources and effectiveness of search terms selected during mediated interactive information retrieval. The study focuses on determining the retrieval effectiveness of search terms identified by users and intermediaries from retrieved items during term relevance feedback. Results show that terms selected from particular database fields of retrieved items during term relevance feedback (TRF) were more effective than search terms from the intermediary, database thesauri or users' domain knowledge during the interaction, but not as effective as terms from the users' written question statements. Implications for the design and testing of automatic relevance feedback techniques that place greater emphasis on these sources and the practice of database searching are also discussed.

POSTCO human-computer interaction
 information retrieval
 online information systems
 searching
 user studies

PRECO human-computer interaction
 information retrieval--searching
 online information systems--user studies

PUBLICATION Information Processing and Management
YEAR 1995

DOC_NO 15
AUTHOR Wang, Peiling & Soergel, Dagobert
TITLE Beyond Topical Relevance: Document Selection Behavior of Real Users of IR Systems
CITE v. 30, October 23-28, 87-92

ABSTRACT This paper reports on part of a study of real users' behavior in selecting documents from a list of citations resulting from a search of an information retrieval (IR) system. Document selection involves value judgments and decision making. Understanding how users evaluate documents and make decisions provides a basis for designing intelligent IR systems that can do a better job of predicting usefulness. Twenty-five faculty and graduate students in an academic department participated in the study with requests presented to the department's library. After a reference interview to establish the user's profile (information needs, tasks at hand, expectations, and experience), a search was done on Dialog. From the resulting printouts of full citations participants selected the documents for which they wished to see the full text. Participants were asked to read aloud (information attended) and think aloud (whatever comes to their minds, decisions and their reasons). Analysis of the verbal protocol data identified the criteria, the sources of information, and the patterns of document selection behavior. Among the factors identified are topical relevance, field relevance (field is broader than topic), novelty, expected usefulness (from previous experience of the author or journal), orientation, recency, availability, special requisites and time constraints. The design of an intelligent IR system requires an enormous knowledge base including not only the domain knowledge, but also the individual user's preferences and document selection behavior. Such a system could better anticipate the user's decisions and present its output accordingly.

POSTCO document selection
information retrieval
information seeking behavior
online information systems
user studies

PRECO information retrieval--document selection
information seeking behavior
online information systems--user studies

PUBLICATION ASIS '93: Proceedings of the 56th ASIS Annual Meeting: Integrating Technologies; Converging Professions

YEAR 1993

References

Meadow, Charles; Boyce, Bert, and Kraft, Donald. (2000). Text Information Retrieval Systems, 2nd ed. San Diego: Academy Press, 276-277.

Records are well structured. Vocabularies are assigned logically, and the 2 vocabulary fields are parallel.

PART B

Evaluation

Criteria

We decided that in order to be termed a success, our tests needed to meet a threshold of 20% recall and 40% precision. We felt that given the small size of the database and the limited size of the vocabulary used in search strings, this was an acceptable level.

Where to search?

It was decided that searches would be conducted in one natural language field, the abstract, and in the two controlled-vocabulary fields, the pre-coordinate and post-coordinate vocabularies. Since the abstract field often contains a wealth of information about the article, we felt that this made a logical choice for our natural language searches. It seemed especially appropriate because it is a reasonable first stop for many searchers who have little to no knowledge of the inner workings of information retrieval systems.

Search strategies and testing procedure

In order to conduct our tests, we had to place ourselves in the shoes of a potential searcher. That meant having to pick things to search for in the database, which further meant the construction of questions. Our group came up with 14 questions that we felt a student might use in compiling information for a paper or perhaps even just to help clarify topics discussed in class. They covered a wide range of topics related to information retrieval, from Boolean logic's role in information retrieval to the definition of an online information system. In order to determine which documents we felt were relevant to our questions, we looked both at the articles' abstracts as well as their pre-

coordinate and post-coordinate vocabularies. While user-determined relevance is far from perfect, it is exactly the procedure a database user would employ. [Good points.](#)

We examined our questions critically and isolated the key concepts covered by each one. The concepts were translated into search terms for each search field – the natural-language field and the two controlled-vocabulary fields. For the natural language field, we simply utilized the concepts we had pulled from the questions. They were then used as search strings utilizing truncation and Boolean operators, specifically AND and OR and modified when necessary. Searching the vocabulary fields was done in much the same way; however, as we refined our searches in these fields, searching entailed pulling concepts from our validation lists that we felt most closely approximated our search need and inserting them as search strings with Boolean operators. [Good. Controlled vocabularies always work best if you can search for the terms that are contained in them.](#)

As each test was completed, we made a record of our search string and the documents that were retrieved with each search. We utilized the formulas given for recall and precision to gauge each search's level of each measure. We then continued to modify our searches until the level of precision and recall we sought was achieved. This mirrors the average user, who would continue to refine the search until he found what he was looking for [good point \(although unfortunately the research shows that users actually give up after one try\)](#), though he would most likely not utilize mathematical measures of recall and precision, instead relying on his own mental measurement of relevance.

Tests

An Example Test Search

1. Created question #3. What influences a user's search of an online system?

2. Came-up with the concepts we will search:
user, searching, behaviors, online systems

3. Thought that docs. 2,6,8,9,10,11,14,15 are relevant.
Total: 8 in the original set.

4. Abstract search strings:

1st attempt: user* & search* & behavior* & online system*. No hits. too specific.

2nd: search* & system* & strateg* 2 hits. 15,9. Too limited.

3rd: search* & system* 7 hits 1,2,9,11,12,14,15.

Recall=number of docs retrieved in original set / total number of docs in original set $5/8 = 62\%$.

Precision=number of docs retrieved in original set / total number. of docs. retrieved $5/7 = 71\%$.

5. Post co search strings:

1st attempt: same as in abstract user* & search* & behavior* & online system*. No hits.

2nd: search* & strateg* & online system*. No hits, too specific Try F3.

3rd: search* & document relevance / =search terms & =online information systems. 1 hit 1, too limited.

4th: search* & document relevance. 3 hits, 5,6,12, 1 in original set.

Recall= $1/8 = 13\%$.

Precision= $1/3 = 33\%$.

6. Pre co search string:

1st attempt: search* & online information system*. No hits, use the F3 key.

2nd: information retrieval search* & online information systems user stud*. 2 hits, 8,14, both in original set.

Recall= $2/8 = 25\%$.

Precision= $2/2 = 100\%$.

3rd: Modify search to include something about human behavior, like online information systems human factors / human computer interaction / information retrieval search*. 3 hits, 7,11,14, 2 in original set.

Recall= $2/8=25\%$.

Precision= $2/3= 67\%$.

4th: online information systems human factors & information retrieval search*/ human computer interaction & *information retrieval search* 3 hits, 7,11,14, 2 in original set.

Recall and precision same as above.

[Your approach to the search and strategies is very logical in these 3 groups of searches.](#)

Analysis

Since most of our tests achieved results that met or exceeded our threshold for both precision and recall, we believe our database and vocabulary are successful

instruments for their intended purpose. Our work, however, showed that our procedures and vocabulary were far from perfect. When we first ran our searches, we blithely began taking our concepts which we culled from the questions and entering them as single-term searches instead of combining terms using Boolean operators to include the multiple concepts in our search questions. Though we eventually caught our mistake and were able to correctly execute searches, we wanted to be aware of what had lulled us into such **potentially reckless** 😊 action. While we were at a loss to precisely explain why such action was taken, we believe we briefly became fixated on the idea of term indexing and believed that using the search terms would allow us to pull up what we wanted. It did retrieve many of the documents we sought, but our rates of recall and precision were severely skewed. [Good for you, for catching this!](#)

We feel, in addition, that as a search tool, a post-coordinate vocabulary is much more useful to a searcher, as it allows a greater degree of freedom when creating his or her search. They are not constrained by the rigidity that is inherent in a pre-coordinate vocabulary such as the Library of Congress Subject Headings. Individuals tend to think in smaller chunks of information, which are better represented by post-coordinate vocabularies. We also felt that our post-coordinate vocabulary terms should have been broken down into smaller parts, such as: systems, online, users, strategies, information, so that users could search. We decided that in a database limited to a single field, information retrieval, a complete lexicon of terms pertaining to information retrieval should be available in the post coordinate vocabulary. It is not necessary that all terms will apply to the current list of articles, but if the database is to grow, the vocabulary should aim to include a complete lexicon for information retrieval.

Well, yes, but you don't want to have terms for users to search on that won't retrieve a single document. You could create such a complete list, but only add terms to the validation list when they were actually assigned to articles.

We discovered other problems with our pre-coordinate vocabulary. We relied too heavily on assigning second-tier terms instead of developing them into top-level terms as well, which would have allowed for a more general representation of the articles' contents. I think I understand what you mean here, and you're correct, as long as it doesn't result in having, for instance, both a term for *online information systems—searching* and *searching—online information systems*. In addition, our pre-coordinate vocabulary was completed from the standpoint of having the broadest term then broken into a smaller term, like “online information systems—searching.” After testing we have realized that searchers don't always think in a top down fashion, especially students who are trying to build a bridge between their own ideas and the concepts of others. Moving from the specific to the abstract in a bottom up fashion is more in tune with user warrant, whereas the top down, inverse word order, is more in tune with literary warrant. Literary warrant makes more sense to the vocabulary developers and the indexers. But user warrant of course makes more sense to the user. Hm. I'm still not understanding what you mean by inverse word order. I think the key issue here is actually which part of the term is considered to be the most fundamental to the discipline or the user. Users may see searching as the fundamental thing they're interested in and want to know what different kinds of searching they can find information about – searching online systems, searching manual systems, searching the internet, etc.; then, as you say, *searching* as the first part of the term and *online systems* as the subdivision makes sense. Within the field, searching may also be seen as the fundamental topic of interest, with what is being searched of secondary importance. In that case, literary warrant also justifies making *searching* the subject and *online systems* the subdivision. If “the field” considers the type

of system the most fundamental concept of interest and what's being done with it (searching, designing, evaluating) the secondary consideration, then that's an argument for *online systems* being the subject and the other 3 being the subdivisions.

Thus we reconsidered our initial design strategy and decided we may have a better representation of the concepts in the articles and a more useful database if we had not used inverted language in the pre-coordinate vocabulary and instead use natural language subject headings. Thus, the string would be “searching – information retrieval,” instead of “information retrieval—searching,” a change we felt would make it easier for the user to find the term she/he thinks is most useful for their search, since most users approach a search with inductive reasoning. Top down is deductive reasoning, where you start with the rule and deduce the example. Bottom up inductive reasoning, you start with the example and induce the rule. *This is an interesting approach to the problem; I'll continue to think about it.*

If you were writing scope notes for your vocabulary, how would you differentiate between *searching—information retrieval*, and *searching* by itself? When would you assign one and when the other?

We found that our choices of pre-coordinate vocabulary were too specific, too limited, and did not allow for enough general, hit-the-side-of-the-barn type, searches. We decided in the design phase to remove the pre-coordinate main headings as stand-alone terms, instead insisting that they must have a second-tier term in order to appear in our vocabulary. This made the pre-coordinate vocabulary field less flexible and less useful for searching broad, general topics; though it may have increasing precision it decreased recall. *It wouldn't actually have done anything to precision and recall, if the indexer assigned whichever term (stand-alone or with subdivision) best reflected the scope of the article, and if the searcher had chosen the term that best expressed his/her information*

need. But it would indeed have made the field more flexible, since then those choices would have been available to the indexer and the searcher.

If we had more time, we would have gone back to the database fields and changed their names to ones that would be more representative of their contents and more recognizable by users. Abstract could have been called Keyword searches, Postco renamed as Subjects, and Preco, Subject Headings, names that would tell users what to expect and how to search. *That's a useful idea. One thing I've learned in database design is that field names are much more important than I would have thought – they really make things quickly understandable .. or they don't.*

We found DBTextWorks to be a cumbersome tool for searching a database and found it surprising that some small libraries actually rely on the program to form their online catalogs. For instance, when searching the pre-coordinate vocabulary validation list, picking out appropriate terms requires you to examine and reexamine the list because of the strange way that DBTextWorks removes all punctuation eliminating any distinction between main heading and subheading and turning them into a single long subject heading that clearly represents more than one concept. *Yes, this is a problem. DBT works better with postcoordinate vocabularies than with precoordinate.* This seems to defeat the purpose of constructing subject headings in a two-tiered format to begin with and is mind boggling to the user. While the program does a good job of executing searches, we found it an unfriendly platform. While we understand that DBTextWorks is used for a specific reason within this class and assignment, we also realize that a truly useful database would need to be built with a more flexible, user-friendly interface. *It*

could certainly have a simpler interface ... but it would be hard to find one that gave more flexibility and power to the searcher.

Another problem with DBTextWorks is the truncation issue. You can only truncate the final word in a term (phrase). If, in the pre-coordinate fields, you truncate the last word of the main heading, your search comes up empty. [This isn't right; there's some misunderstanding here about how to use the truncation feature in DBT.](#) (Much) later comment: I finally figured this out as I was looking at your searches. I think what you're referring to here is the fact that DBT can't do left-truncation. When you search on *system** in an attempt to retrieve online information systems, you are essentially asking the database to do left-truncation (ie., truncate *online information* and retrieve any descriptor that contains the word *system**). Remember that we required that you term index this field; this means that the entire term is in there as a single unit. Almost as if it were *onlineinformationsystems* (except it does read the spaces). And just as you can't do left truncation to search for (for example) **cycle* to retrieve both bicycle and unicycle, you can't ask it for search for the equivalent of **system**. If we had allowed you to word index the field, you would be able to do this, because each word would have been a unit in the inverted file. However, if the field had been word indexed, you'd have lost the meaning that is inherent in the term "online information systems." The problem arises from the fact that we made you term index the field to preserve the terms as terms, not from a quirk in how DBT does truncation. So you can't truncate down to the root any but the last word of a term. Because natural language fields allow both word and term searches, you can truncate individual words anywhere in the search string. Because you can do this in abstract, users might think they can do this in every subject access field. So

we should have warned them of this by including it in the User Guide's rules for searching section.

Discussion and personal evaluations

We made some big mistakes with our initial test searches. Instead of combining concepts in a single search string, we searched each concept individually and then worked out the recall and precision rates for documents retrieved based on these single word searches. This made no sense because the documents in our original sets were chosen based on their containing all the concepts in the question, not each concept individually. Only by combining concepts from the test question and then searching the database using search strings of combined concepts would we be testing whether or not the database found the set of relevant documents we were aiming for.

We clearly did not understand that our goal for evaluation was to compare the natural language fields of **title** *I don't actually see that you did any testing in the title field.* and abstract, with the controlled vocabulary fields of pre-coordinate and post-coordinate. Natural language searches are keyword searches, so combining keywords is essential to retrieving articles relevant to the concepts in the original question. Pre-coordinate vocabulary is by definition a combined vocabulary, one that combines main subject headings with subheadings. Because we wanted to compare results of post-coordinate searches with pre-coordinate search and with natural language searches, we would have to **combine post coordinate terms and abstract search terms**, *I don't understand this. It seems that you would combine post coordinate terms with a Boolean AND. For instance, if you had a pre-coordinate term for *reference services—evaluation*, in the postco field you would search for *reference services & evaluation*.* since the pre-

coordinate terms are already combined. We had to compare like with like. So we ended up running two sets of tests. Our original set results were meaningless because the search strings did not adequately represent the users' multi-concept questions, requiring combined term searches. And this database does not allow combined set searching. [If you mean that it doesn't allow you to retain previously created sets as Dialog does, you're right. It would be great if it did, but I don't know of any general purpose dbms software that does allow this \(which doesn't mean there aren't any, of course\).](#)

In order to carry out a more effective testing phase, we decided that, ideally, a fourth group member with the role of user/tester would have been beneficial. [Yes.](#) This member would be able to provide the kind of feedback usually gained in a requirements elicitation study where the needs of users are studied before the design phase starts. As noted earlier, we may have been [let](#) to design our pre-coordinate vocabulary differently. One team member ran into difficulty when developing his pre-coordinate vocabulary because he had difficulty grasping the ideas presented by hierarchical, top down, design of our vocabulary. He felt more comfortable with natural language design where the more specific idea are presented first, and the broader context presented second. To use the example presented earlier, this member would have found “searching—online information systems, searching—library catalogs, searching—subject headings” more useful than the form we used in our vocabulary, “online information systems—searching, library catalogs—searching.”

Conclusion

Despite our frustrations, we all agree that this assignment is outstanding in providing us an opportunity to design a vocabulary and build a database. We come away

with a respect for the retrieval systems that consider the psychology of the user in their design. Meadow, et al., state that utility is the ultimate measure of how well the results of a search satisfy users needs. “If a user finds it easy to conduct the search he or she wants and is able to quickly and easily retrieve all and only the relevant documents to his or her query, then the database has served its purpose” (317). We believe that with further refinement and construction, our database could become one that is rated with a high level of utility by users. We have discovered more useful ways to construct our vocabularies and are more easily able to envision the search process from the user’s point of view.

References

Meadow, C., Boyce, B., & Kraft, D. (2000). Measurement and Evaluation. *Text Information Retrieval Systems*. (2nd ed.). San Diego: Academic Press, 310-337.

Test results

Q1. What affects the design of library catalogs?

7 in original set, 1,7,9,11,12,13,14.

Abstract: design* & library catalog*. 0 hits. So try design* & catalog*, 0 hits.

2nd try: design* & system*. 8 hits, 1,5,4,10,11,12,14,15, 4 in original set. Recall = $4/7 = 57\%$, precision = $4/8 = 50\%$.

Post co: design* & system* nothing so try, design* & online info sys you get 4 hits, 1,9,10,11. 3 are in original set. Recall= $3/7 = 43\%$, precision= $3/4 = 75\%$.

Pre co: online information systems design* 5 hits, 1,9,10,11,12. 4 were in original set. Recall= $4/7 = 57\%$, precision = $4/5 = 80\%$.

Q2. What are online information systems?

Key concepts: online information systems, library systems. 3 in original set, 7,11,13.

Abstract: information system*/ library system*/online system* 3 hits, 9,10,11. 1 is in original set. Recall= $1/3=33\%$, precision= $1/3, 33$

Post co: system* 0 hits. Try: information system* 0 hits.

2nd: use F3 key for controlled vocabulary: online information systems. 9 hits, 1,7,8,9,10,11,13,14,15. 3 were in original set. Recall = $3/3 = 100\%$, precision $3/9, = 33\%$.

3rd: theories/=history & =information processing &=online information systems. 1 hit, 9, not in the original set. Recall 0, precision 0.

4th: theories & =information processing &=online information systems 1 hit, 9, not in original set. Recall 0, precision 0.

5th: =online information systems & history 1 hit, 13. 1 in original set. Recall= $1/3 = 33\%$, Precision $1/1= 100\%$ [It seems you got away from your question here – where did the concepts of history and information processing come in?](#)

Pre co: try information systems* 0 hits, so use F3 key.

Try a general search with OR online information systems history / =online information systems user studies / =online information systems design. 9 hits, 1,8,9,10,11,12,13,14,15. 2 were in original set. Recall = $2/3 = 66\%$, precision = $2/9 = 22\%$.

Q3. What influences a user's search of an online system.

8 in original set, 2,6,8,9,10,11,14,15.

Key concepts: users, searching behaviors, online systems.

Abstract: user* & search* & behavior* & online system*. 0 hits.

2nd: search* & strateg* & system*. 2 hits, 9,15, all in original set. Recall $2/8=25\%$, precision $2/2=100\%$.

3rd: search* & system*. 7 hits, 5 in original set. Recall= $5/8=63\%$. Precision= $5/7=71\%$.

Postco search string: user* & search* & information system* came up 0. go to F3.

change search to search* & online information system*. 5 hits, 3 in original set.

Recall= $3/8=38\%$, precision= $3/5=43\%$.

Preco search: try search* & online information system*. No hits.

Try F3, information retrieval search* & online information systems user stud*. Results: 2 hits.

2nd: online information systems human factors / human computer interaction / information retrieval search*.

3rd: online information systems human factors & information retrieval search*/ human computer interaction &*information retrieval search* online information systems human factors & information retrieval search*/human computer interaction &*information retrieval search*

Results: 3 hits, 2 in original set. Recall= $2/8= 25\%$, precision is $2/3=67\%$.

Q4. How will digital documents be preserved for the future?

1 in original set, 3. Key concepts: digital document*, future, preserv*.

Abstract search string: digital doc* & future*. Results 1 hit, 3.

2nd: electr* & doc* & preserv* 1 hit, 3. Recall 1/1=100%, precision 1/1=100%.

Post co search: digital doc* & future* 0 hits

2nd: digital doc* & preserv* 1 hit, 3. Recall 1/1=100%, precision 1/1=100%.

Pre co: digital documents preserv* 1 hit, 3. Recall 1/1=100%. precision 1/1=100%.

Q5. How can a librarian effectively assess a user's needs?

3 in original set, 2,8,10. Key concepts: users, needs, librar*, interview*.

Abstract search string: reference* & user* & question*/need*

Results 4 records, 2,6,10,15. 2 in original set. Recall=2/3=67%, precision=2/4=50%.

Post co search string: reference* & user* & question*/need*

Results 3 records, 2,8,10. All in original set. Recall 100%, precision 100%.

Pre co search string: =online information systems user stud* / =reference interviews questioning strategies.

Results 5 hits, 2,8,10,14,15, 3 in original set. Recall=3/3=100%.
precision=3/5=60%.

Q6. What leads a user to select specific documents?

Relevant documents: 6,8,10,15 (4 documents).

Key concepts: users, document selection, relevance judgements

Abstract: user* AND relevance (1 retrieved, 1 in original set, 1/4=25% recall, 1/1=100% precision)

Post co: document selection AND information-seeking behavior OR user studies 6
retrieved: 1,5,8,10,14,15. 3 in original set, 3/4= 75% recall, 3/6= 50% precision)

Pre co: information retrieval – document selection AND online information systems –
user studies OR information seeking behavior (3 retrieved: 6,8,15. 3 in original set, 3/4 =
75% recall; 3/3=100% precision))

Q7. What influences the creation of a controlled vocabulary? Original set 1,5,7,13, 4 documents.

Key concepts: vocabulary, design, information systems.

Abstract: vocabulary & design & information systems, 0 hits.

2nd: search term* & design & system* 1 hit, 14. Recall=1/4=25%, precision 1/1=100%.

Post co: search term* & design & system*. 0 hits.

2nd: search term* & design & online information system*. 1 hit, 1.

3rd: search term* & design & online information system* / =access points / =descriptors / =digital documents / =thesauri. 4 hits, 1,3,5,8, 2 in original set. Recall 2/4 = 50%, precision 2/2=100%.

Pre co: search term* & online information systems design*/ =access points / =descriptors / =digital documents / =thesauri. 1 hit, 1.

2nd: search term* / online information systems design*/ =access points / =descriptors / =digital documents / =thesauri. 6 hits 1,5,9,10,11,12, 2 in original set. Recall=2/4=50%, precision 2/6=33%.

Q8. What are the similarities and differences between computerized and human approaches to information retrieval? Original set 4,6,7,9,11,12,13,14, 8 documents.

Key terms: computer, human, information retrieval.

Abstract: computer* & human* & information retrieval* 0 hits.

interaction & computer* 0 hits.

2nd: system* & human* & information retrieval* 1 hit, 14.

3rd: information retrieval & stud*. 2 hits, 14,15.

4th: information retrieval & us* & stud*. 2 hits, 14,15. Recall 1/8=13%, precision=1/2=50%.

Post co: information retrieval & user stud*. 5 hits 1,5,8,14,15, 1 in original set. Recall 1/8=13%, precision 1/5=20%.

2nd: information retrieval & human-computer interaction. 2 hits, 7,14. Recall 2/8= 25%, precision 2/2=100%.

Pre co: information retrieval & human-computer interaction, 0 hits.

2nd: =information retrieval human factors / =information processing human cognition / =human computer interaction / =information retrieval user studies / =information retrieval theories / =online information systems human factors. 9 records, 2,4,6,7,9,10,11,13,14, 7 in original set. Recall 7/8=86% precision 7/9=78%.

Q9. What is Boolean logic's role in information retrieval?

Relevant documents: 13 (1 document).

Key concepts: Boolean logic.

Abstract: Boole* (#13 retrieved) (1 retrieved, 1 relevant) ($1/1=100\%$ recall, precision)

Post co : Boolean algebra (#13 retrieved) (1 retrieved, 1 relevant) ($1/1=100\%$ recall, precision).

Pre co: information retrieval—Boolean algebra (#13 retrieved) (1 retrieved, 1 relevant) ($1/1=100\%$ recall, precision).

Q10. What factors have influenced the evolution of information retrieval? Relevant documents: 4,7,9,11,12,13 (6 documents)

Key concepts: information retrieval, history, theories, design.

Abstract: theories OR design (12 retrieved: 1,4,5,6,8,9,10,11,12,13, 14,15). (5 relevant. $5/6=83\%$ recall; $5/12=42\%$ precision)

Post co: information retrieval AND history OR theories OR design (1,4,6,9,11,12,13 retrieved) (7 retrieved, 5 relevant. $5/6=83\%$ recall; $5/7=71\%$ precision)

Pre co: information retrieval—theories OR online information systems—design OR online information systems—history. (1,6,9,10,11,12,13 retrieved) (7 retrieved, 4 relevant. $4/6=66\%$ recall, $4/7=57\%$ precision).

Q11: How are users' needs considered in the design of information systems?

7 in original set, 4, 6, 7, 8, 10, 14, 15.

Abstract: user* & need*, 2 hits, 10,15. 2 are in the original set. Recall is $2/7=29\%$ and precision is $2/2=100\%$. Then tried user* & design*, 5 hits, 1, 9, 10, 14, 15. 3 are in the original set. Recall is $3/7=43\%$ and precision is $3/5=60\%$.

Post-co: online information systems, 10 hits, 1, 5, 7, 8, 9, 10, 11, 13, 14, 15. 5 are in the original set. Recall is $5/7=71\%$ and precision is $5/10=50\%$

Pre-co: information seeking behavior, 3 hits, 6, 8, 15. 3 are in the original set. Recall is $3/7=43\%$ and precision is $3/3=100\%$. Then tried information seeking behavior / online information system--design, 8 hits, 1, 6, 8, 9, 10, 11, 12, 15. 4 are in the original set. Recall is $4/7=57\%$ and precision is $4/8=50\%$.

Q12: What are some school of thought regarding the design of information systems?
7 in original set, 1, 7, 9, 11, 12, 13, 14.

Abstract: information system*, 2 hits, 10, 11. 1 is in the original set. Recall is $1/7=14\%$ and precision is $1/2=50\%$. Then tried design*, 8 hits, 1, 5, 9, 10, 11, 12, 14, 15. 5 are in the original set. Recall is $5/7=71\%$ and precision is $5/8=63\%$.

Post-co: design, 5 hits, 1, 9, 10, 11, 12. 4 are in the original set. Recall is $4/7=57\%$ and precision is $4/5=80\%$. Then tried online information systems, 10 hits, 1, 5, 7, 8, 9, 10, 11, 13, 14, 15. 6 are in the original set. Recall is $6/7=86\%$ and precision is $6/10=60\%$.

Pre-co: information retrieval--theories, 2 hits, 6, 13. 1 is in the original set. Recall is $1/7=14\%$ and precision is $1/2=50\%$. Then tried online information system—design, 5 hits, 1, 9, 10, 11, 12. 4 are in the original set. Recall is $4/7=57\%$ and precision is $4/5=80\%$.

Q13: What role does psychology play in information retrieval?
7 in original set, 4, 6, 8, 9, 10, 11, 15.

Abstract: psycholog*, 2 hits, 6, 9. 2 are in the original set. Recall is $2/7=29\%$ and precision is $2/2=100\%$. Then tried psycholog* / information retrieval, 6 hits, 5, 6, 7, 9, 14, 15. 3 are in the original set. Recall is $3/7=43\%$ and precision is $3/6=50\%$.

Post-co: information seeking behavior, 3 hits, 6, 8, 15. 3 are in the original set. Recall is $3/7=43\%$ and precision is $3/3=100\%$. Then tried information retrieval, 11 hits, 1, 2, 5, 6, 7, 8, 11, 12, 13, 14, 15. 4 are in the original set. Recall is $4/7=57\%$ and precision is $4/11=36\%$.

Pre-co: psychology--theories, 3 hits, 4, 6, 9. 3 is in the original set. Recall is $3/7=43\%$ and precision is $3/3=100\%$. Then psychology—theories / information seeking behavior, 5 hits, 4, 6, 8, 9, 15. 5 are in the original set. Recall is $5/7=71\%$ and precision is $5/5=100\%$.

Q14: How do human beings process information?
5 in original set, 4, 7, 9, 11, 15.

Abstract: human*, 5 hits, 4, 9, 11, 12, 14. 3 are in the original set. Recall is $3/5=60\%$ and precision is $3/5=60\%$. Then tried human* & information*, 4 hits, 4, 9, 11, 14. 3 are in the original set. Recall is $3/5=60\%$ and precision is $3/4=75\%$.

Post-co: human cognition, 2 hits, 4, 9. 2 are in the original set. Recall is $2/5=40\%$ and precision is $2/2=100\%$. Then tried human cognition / human computer interaction, 4 hits, 4, 7, 9, 14. 3 are in the original set. Recall is $3/5=60\%$ and precision is $3/4=75\%$.

Pre-co: human computer interaction, 2 hits, 7, 14. 1 is in the original set. Recall is $1/5=20\%$ and precision is $1/2=50\%$. Then information processing—human cognition, 2 hits, 4, 9. 2 are in the original set. Recall is $2/5=40\%$ and precision is $2/2=100\%$.

Evaluation methodology is very carefully thought-out and well designed. One serious problem is that I can't see that any evaluation was done of one of the natural language fields, the title field.

Discussion shows a strong grasp of concepts, issues, and goals in information retrieval.

Users guide is very well done.