
Retrieval features for online databases: common, unique, and expected

*Roslina Othman and
Nor Sahlawaty Halim*

The authors

Roslina Othman is an Assistant Professor and **Nor Sahlawaty Halim** is an MLIS Graduate, both at the Department of Library and Information Science, Faculty of Information and Communication Technology, International Islamic University Malaysia, Kuala Lumpur, Malaysia

Keywords

Online databases, Information retrieval

Abstract

The aims of this study were to identify the retrieval features for online databases; difficulties faced by users; and retrieval features expected by users. A total of 25 databases were surveyed and 40 users were interviewed after the training sessions. Common retrieval features included Boolean operators, phrase searching, match of exact words or phrases, field specific and limit fields searches, truncation, and wildcard. Even though features are offered in many systems, their interpretation and implementation are different. Unique features included lateral searching, density and frequency of terms, reference link, and searching via table of content. The expected features included relevance feedback and term weighting other than those already offered by ACM Digital Library and IEEE Xplore. Such expectations were influenced by the users' background in ICT. Difficulties included application of the retrieval features in searching. Database providers must include the expected features, synonyms linked to terms in the thesaurus, and extensive search examples.

Electronic access

The Emerald Research Register for this journal is available at

www.emeraldinsight.com/researchregister

The current issue and full text archive of this journal is available at

www.emeraldinsight.com/1468-4527.htm

Introduction

Database providers like ProQuest, Ovid, and Cambridge Science Abstracts (CSA) and associations like ACM and IEEE offer systems equipped with retrieval features to help users find relevant items stored in their databases. Retrieval features defined here include those features that help users construct their queries. For example, some have online tutorials that guide users through the features available; basic and advanced searches using Boolean operators, proximity, lateral searching, query-by-example, and re-run saved search strategies. Some retrieval features are common among the database providers, and some are unique.

Common retrieval features are defined here as the features offered and made available by five or more database providers; while unique features are those features offered by less than five database providers. Users expect a set of retrieval features to solve their retrieval tasks. In this paper, these are termed as the expected features. It is possible that users may encounter problems related to retrieval even with common features since each piece of information need requires different approaches to conduct searches. Since all these possibilities are just predictions made based on a review of the relevant literature, it is important to conduct a study that examines the retrieval features, problems faced by users in applying these features to their retrieval tasks, and features that help them retrieve needed items.

The aim of this study was to examine the retrieval features as offered and made available by database providers. The objectives were to identify the retrieval features offered by selected database providers; difficulties faced in using the features; and the features expected by users and missed by the database providers. This study also includes users' first impressions of the retrieval features after a three-month training session and their perception after use.

This study intends to answer the following questions:

- What are the retrieval features that are common to all databases?
- What are the retrieval features that are unique to the databases?
- How are these retrieval features used in the database?
- What are the difficulties faced by these users in applying the features to complete their retrieval tasks?
- What are the retrieval features expected by these users and missed in the database?

Refereed article received 22 December 2003

Accepted for publication 12 March 2004

Literature review

Database providers offer retrieval capabilities in their systems to enable users to retrieve needed items in the collections, be they merely records consisting of only citations or together with full-text. However, as noted by Shneiderman *et al.* (1998), the current user interfaces are confusing. In addition, systems do not provide interpretation of the retrieval features, and there is a lack of standardization in the text-with-embedded-operators among the systems. Even similar databases offered by different providers will have differences in their retrieval features (Tenopir and Hover, 1993). Thus users will most likely fail to find relevant items when they move from one system to another. Shneiderman *et al.* (1998) proposed that the user interface design should have a standard structure and terminology for searching while maintaining the distinct features of individual collections and search mechanisms.

Users should consult several databases that complement each other in terms of content and coverage (Schaffer, 2001). This conclusion was reached based on a study conducted by Schaffer in which he compared nine databases appropriate for political science research in terms of coverage, content, and their simple search features, i.e. AND, adjacency, and truncation. Since each database provider offers different features, there is, thus, a need to identify the features that are available in many systems and those that are not and their implementation to ensure that users will be able to search in more than one database.

Retrieval features offered by most database providers are Boolean operators (Eastman and Jansen, 2003), as well as keyword, phrase, and field searches (Diercks, 2003). Eastman and Jansen (2003) conducted a study to examine the effect of query operators on search results in terms of coverage, relative precision, and relevance ranking. They found that the advanced search screens for the Web search engines did not directly support a full range of Boolean queries, including queries using both AND and OR.

Searching behaviours, information retrieval interfaces and term selection and query expansion have stressed the necessity of some form of thesaurus support at the interface level to help users in selecting search terms for their search (Shiri *et al.*, 2002). Features most often used for thesaurus-enhanced search included: linking search terms with mapping techniques; display of number of documents indexed by the thesaurus term; and look-up and browse options.

Other features expressed as needed by users included: concise and easy-to-follow search tips

and help screens, spelling variations, table of contents (Electronic Collections Committee, 1993); and automatically saved searches and search history (Heidig and Prior, 1997).

Much of the research conducted has focused on the interfaces needed to help beginners in searching (Brajnik *et al.*, 2002; Sutcliffe *et al.*, 2000; Wolfram and Xie, 2002); user satisfaction and usefulness of the system (Ma, 2002); and users' problems in searching (Brajnik *et al.*, 2002; Robins, 2000). Basically, all these studies outlined the different levels of users: beginners who require basic search interface and advanced users. With proper training and practice, the majority of users now no longer belong to the "beginners" category. Searching problems as covered in the literature are mostly relevant for those without training or knowledge about the system.

Issues that are associated with retrieval features of online databases were related to a selection of innovative search tools that attempt to solve specific problems in searching (Kline, 2002). Kline suggested that current search tools exhibit a number of significant weaknesses and users have difficulty in constructing the sophisticated queries needed for effective retrieval. The problems were grouped into four categories:

- (1) searching difficulties;
- (2) retrieval issues;
- (3) document discrimination problems; and
- (4) interface design quandaries.

The importance of providing strategic help menus in an information retrieval system is due to the fact that users do not know how to react in critical situations, and they often do not even realize that their difficulties are due to strategic problems (Brajnik *et al.*, 2002). Other than supplying users with help menus, training and advisor facilities are also found to be essential for effective search strategies and need to be incorporated to enhance the effectiveness of the retrieval features for information retrieval (Sutcliffe *et al.*, 2000).

Shneiderman *et al.* (1998) suggested that systems should also provide relevance feedback, support for successive queries, and keep track of searches in the history buffer to allow review, alteration, and resubmission of earlier searches. Xie (2003) and Xie and Cool (2000) recommended that retrieval features must allow user control. Their studies revealed that users preferred features like a browsing feature for term index, expand, table of content, relevance-feedback, search history. In addition, help mechanisms must be sensitive to technical problems and well organized.

Methodology

This study explored the retrieval features of a total of 25 online databases from 12 database providers (see Table I) using the Help, Guide, or Tutorial menu and trial searches. These databases are currently subscribed to by the International Islamic University Malaysia. However, CSA, the provider for LISANet has to be omitted due to the problem of access during the course of this study. This study involved interviews with a total of 40 users (25 undergraduates, 12 masters students, and three PhD candidates) having a subject background in information and communication technology (ICT).

Stages of the data collection were as follows:

- (1) Survey of the retrieval features for the 25 databases by the researchers.
- (2) A training session that lasted for three months (15 slots), involving lecture, discussion, and hands-on practice, was given to the users.
- (3) Users were interviewed after the training session.
- (4) Users were also interviewed after they accomplished their retrieval tasks using the systems. Their retrieval tasks involved searching assignments on ICT and ICT-related topics, e.g. breast cancer patient

Table I The 12 database providers and the 25 databases surveyed in this study

	Database provider	Databases
1.	Ovid	Medline
2.	ProQuest	ABI/Inform Global Health and Medical Complete Medical Library Psychology Journals Social Science Plus Eric plus Text Education Journals
3.	SilverPlatter	Sociological Abstract EconLit
4.	EI Village	Compendex Inspec
5.	Gale	Legal Trac Expanded Academic ASAP
6.	Biblioline	PsycArticles PsycInfo AIDSearch Child Abuse, Child Welfare and Adoption Info-ASEAN and Pacific Rim
7.	CSA	LISANet
8.	EbscoHost	ERIC
9.	ScienceDirect	ScienceDirect
10.	ACM	ACM Digital Library
11.	IEEE	IEEE Xplore
12.	Emerald	Emerald Fulltext

registry, and which required them to apply all retrieval features identified in this study.

- (5) A control group consisting of 30 users (19 undergraduates, nine master students, and two PhD candidates) who did not undergo the training session were interviewed after they completed the same retrieval tasks using these databases.

Results and findings

Findings from this study are discussed in this sequence: common retrieval features, unique retrieval features, users' first impressions and perceptions after use, difficulties faced by users, and retrieval features expected by these users.

Common retrieval features

Table II shows the retrieval features for each database provider. All systems have two levels of searching: basic and advanced, and provide information on the subject and year covered by the databases. The retrieval features that are available in at least five systems included:

- boolean operators;
- phrase searching;
- match of exact words/phrases;
- field specific searches;
- limit field searches;
- save search;
- search history;
- truncation;
- wildcard;
- have rules of precedence with nested queries;
- proximity search;
- range searching;
- use of thesaurus or permuted index for searching;
- subject search; and
- stemming.

Boolean operators

In all databases, Boolean operators such as AND narrows the scope of a search, OR broadens a search, while NOT eliminates terms from a search.

Boolean operators are grouped with proximity operators (ADJ, NEAR, FREQ, WITHIN) in Ovid and SilverPlatter. In SilverPlatter, Boolean operators are termed as "Query and Set operators". Boolean operators are labelled as "Logical operators" (also called Boolean operators) in IEEE Xplore and in Gale as "Logical operators".

As shown in Table III, Boolean operators are provided at both basic and advanced search levels for some systems, and at advanced search level only for others. Boolean operators are offered in

Table II Retrieval features offered by the 12 database providers

	OVID	CSA	ProQuest	ACM	IEEE	Emerald	EI				Science	
							Village	Gale	Biblioline	SilverPlatter	Ebscohost	Direct
Help menu/online tutorial/guide	X	X	X	X	X	X	X	X	X	X	X	X
Boolean	X	X	X	X	X	X	X	X	X	X	X	X
Rule of precedence	X	X	X	X	X	X	X	X	X	X	X	X
Nested queries	X			X		X		X	X	X	X	
Phrase searching	X	X	X	X	X	X	X	X	X	X	X	X
Exact word/phrase	X	X	X	X	X	X	X	X	X	X	X	X
Stemming	X			X	X	X	X					
Field specific	X	X	X	X	X	X	X	X	X	X	X	X
Limit field	X	X	X	X	X	X	X	X	X	X	X	X
Save search	X	X	X	X		X	X		X	X	X	X
Search history	X	X	X			X	X	X	X	X	X	X
Truncate	X	X	X	X	X	X	X	X	X	X		X
Wildcard	X	X	X	X	X	X	X	X	X	X		X
Proximity	X	X	X	X	X			X	X	X		X
Thesaurus/permuted index	X	X	X		X				X	X	X	
Subject Mapping	X	X	X			X		X				
Range	X	X	X	X	X			X	X	X		
Journal browsing	X				X	X						X
Table of content	X				X	X						X
Classification code												
Explode/expand search	X	X						X				
Lateral searching											X	
Density of terms					X							
Frequency of terms	X											
Reference link	X											
SMARTLinks											X	
Persistent links											X	
Custom links											X	
Sort order	X	X	X	X	X	X		X				
Punctuation marks								X				
Hyphen								X				
Special characters					X							
Treatment type							X					
Automatic translation software											X	
Article types			X									
Query by example				X								
Unary NOT									X			
PIC									X			
Subject authority											X	
Suggest subject headings											X	
Times cited											X	
Spell check											X	

four different ways: pull-down option, entered with the search terms, different box for each operator, and automatic AND operators for multiple terms. It seems surprising that unlike other databases, ProQuest and Emerald provide Boolean operators in the

form of pull-down option at the advanced search level.

Phrase searching and exact words

Phrase search and match of exact words features produce similar results, in which systems must find

Table III Boolean operators

Database provider	Implementation of Boolean operators
1. Ovid	Advanced search – need to be entered with the search terms
2. ProQuest	Advanced search – pull down option
3. Emerald	Advanced search – pull down option
4. SilverPlatter	Search builder – need to be entered with the search terms
5. Ei Village	Basic search – pull down option Advanced search – need to be entered with the search terms
6. Ebscohost	Basic and Advanced search need – to be entered with the search terms
7. ScienceDirect	Basic search – pull down option Quick search – multiple terms will be automatically ANDed Advanced search – need to be entered with the search terms
8. Biblioline	Basic search – pull down option Advanced search – need to be entered with the search terms
9. ACM digital library	Advanced search as desired results – boxes labelled as must have all of the words or phrases, must have any of the words or phrases, must have none of the words or phrases
10. IEEE Xplore	Basic search – pull down option Advanced search – <and >, <or >, <not > and need to be entered with the search terms
11. Gale	Advanced search, subject guide search, and keyword search – need to be entered with the search terms

documents with the exact phrase to appear within the documents or any field as specified before executing the search. The phrase “digital divide” should retrieve documents with the exact terms adjacent to each other within the same document.

However, as shown in Table IV, each database provider allows phrase searching and exact words through different command. ACM Digital Library, Ovid and SilverPlatter treat strings of characters separated by spaces as a phrase; while for Ei Village, Gale and IEEE Xplore, a phrase is any query term enclosed in double quotation marks. Ei Village also accepts braces, { }, for phrase and exact words. The check-box labelled as “Exact Match” in Ebscohost and Emerald, and the “contain phrase” in the pull-down menu in Biblioline, are used for phrase and exact match searches.

Field specific searches

Field specific searches allow users to limit their query terms to a specific field(s), which includes abstract, author, title, accession number, and

Table IV Phrase searching and exact words features

Database provider	Implementation of phrase searching and exact words features
1. Ovid	Space for phrase searching Multiple terms without Boolean operators for phrase search
2. ProQuest	Double quotation marks for phrase search and exact words Multiple terms without Boolean operators for phrase search
3. Emerald	Quick search – check the box next to Phrase for phrase search and check the box next to Exact Match for exact match Advanced search – double quotation marks for phrase search and exact match
4. SilverPlatter	Space for phrase searching Multiple terms without Boolean operators for phrase search
5. Ei Village	Quotation marks or braces { } for phrase and exact words
6. Ebscohost	Words without Boolean operators for phrase searching Check the box next to Exact Match for exact match Quotation mark for exact words with stopwords ignored
7. ScienceDirect	Double quotation marks for phrase search and exact words
8. Biblioline	Quick mode – “contain phrase” in pull down option Advanced mode – typing the phrase in desired search fields Expert mode – typing the phrase in any search set
9. ACM digital library	Space for phrase searching Multiple terms without Boolean operators for phrase search Double quotation marks for exact phrase or words regardless of case Double quotation marks and capitalized first letter for exact phrase or words and upper case for first letter
10. IEEE Xplore	Double quotation marks for exact words or phrase <phrase > command for exact phrase <word > command for exact word
11. Gale	Double quotation marks for exact words or phrase

subjects. Limit search is another feature that constrains a users’ query into a certain requirement; for instance by limiting the search to “full text available” and “English language” documents only. Limiting the search to English language documents is available in all systems.

Truncation and wildcard

Truncation and wildcard are not mentioned in the Help menu of Ebscohost; however, the trial searches showed that truncation can be done in Ebscohost using the symbol (*). Symbols used to represent the truncation and wildcard in systems offered by 11 database providers are shown in Table V. Of the 11 database providers, only five (Ovid, ProQuest, Emerald, ScienceDirect, and SilverPlatter) offer these two features and treat them as two independent commands. Ei Village, Ebscohost, and Biblioline have truncation only, while ACM, IEEE Xplore, and Gale have wildcard only. Truncation and wildcard are not represented by a standard symbol. The symbols (* and ?) are used to represent truncation in some systems and to represent wildcard in others.

In addition to the different symbols, truncation and wildcard are also interpreted differently. For instance, the truncation in ProQuest means a right-hand truncation that will find all forms of a word, while in SilverPlatter it substitutes a string of zero or more characters. The wildcard feature in ProQuest replaces any single character, either within the word or at the right end of the word, while in SilverPlatter it substitutes for one character or more. Gale offers wildcard with the symbol (*) to substitute for any numbers of characters; the symbol (?) for exactly one character; and the symbol (!) to substitute for one or no characters.

Ovid offers unlimited truncation and limited truncation, and mandated wildcard and optional wildcard. Unlimited truncation represented by \$ or “:” will retrieve all possible suffix variations of a root word; while limited truncation requires a digit after the command to specify maximum suffix characters. Mandated wildcard represented by # is a substitute for one required character and useful for some plural forms. Optional wildcard, represented by ? is a substitute for one or no

Table V Truncation and wildcard features

	Database provider	Symbol used for truncation	Symbol used for wildcard
1.	Ovid	\$	# or ?
2.	ProQuest	?	*
3.	Emerald	Check box	*
4.	SilverPlatter	*	?
5.	Ei Village	*	
6.	Ebscohost	*	
7.	ScienceDirect	!	*
8.	Biblioline	* and ?	
9.	ACM digital library		* and ?
10.	IEEE Xplore		* and ?
11.	Gale		*, ? and !

character and useful for British and American variant.

Rules of precedence

A total of 11 database providers offer rules of precedence, and seven of them (Ovid, Ebscohost, ACM Digital Library, Emerald, Gale, Biblioline, and SilverPlatter) have nested queries. ScienceDirect offers a prioritizing search feature that has a similar interpretation to the rules of precedence. The rules of precedence determine the order of execution, either AND, OR, NOT, and any other operators or from left to right or the sequences of the box or pull down option for Boolean operators like Ei Village and IEEE Xplore (see Table VI). Unlike other database providers, ProQuest and ScienceDirect give precedence to OR and Gale gives precedence to NOT operator. The nested queries will execute operators specified in parentheses first before those without the parenthesis.

Proximity search

A proximity search feature is offered by eight database providers. The most common application includes words adjacent to each other in any or in a specified order and *nas* number of words between the search terms. As shown in Table VII, commands used for proximity include ADJ (Ovid, SilverPlatter, and Biblioline), WITH

Table VI Rules of precedence and nested queries

	Database provider	Rules of precedence and nested queries
1.	Ovid	Left to right for words without parenthesis In parenthesis
2.	ProQuest	Left to right OR, AND
3.	Emerald	In parenthesis
4.	SilverPlatter	In parenthesis
5.	Ei Village	Terms in first two boxes followed by the third one
6.	Ebscohost	AND, OR In parenthesis
7.	ScienceDirect	OR, Within (W/nn), Precedes (PRE/nn), Not Within (NOT W/nn), Two words in same field (W/SEG), Not the two words in the same field (NOT W/SEG), AND, AND NOT
8.	Biblioline	AND, OR In parenthesis
9.	ACM digital library	AND, OR In parenthesis
10.	IEEE Xplore	Basic search – terms in the first two boxes followed by the third one Advanced search – in parenthesis
11.	Gale	NOT, AND, OR

Table VII Proximity search feature

	Database provider	Commands
1.	Ovid	adj, freq (frequency)
2.	ProQuest	Within, Within Doc, Not Within, words within a field (PRE)
3.	SilverPlatter	adj, near, with
4.	ScienceDirect	Within (W/nn), Precedes (PRE/nn), Not Within (NOT W/nn), Two words in same field (W/SEG), Not the two words in the same field (NOT W/SEG)
5.	Biblioline	adj, near
6.	ACM digital library	near, sentence, paragraph
7.	IEEE Xplore	near, sentence, paragraph, order
8.	Gale	near, with

and/or WITHIN (ProQuest, Gale, ScienceDirect, and SilverPlatter), and NEAR (ACM Digital Library, Biblioline, Gale, and IEEE Xplore). In the case of Ovid Medline, *n* is up to 99, and in IEEE Xplore, *n* is up to 1,024.

A proximity search used to retrieve words occurring in the same sentence and paragraph is available with ACM Digital Library and IEEE Xplore, while the proximity feature in IEEE Xplore allows search terms to occur in the same order.

Range searching

Range searching is commonly designed for limiting the search to a Publication Year. Biblioline and ACM use <, >, =, >, <= operators in Publication Year field, while Ovid, ProQuest, and IEEE Xplore provide pull down menus or scroll button for users to select the desired year.

Thesaurus

A thesaurus or a permuted index, consisting of synonyms, related terms and preferred terms, is offered by seven database providers. These databases provide a Thesaurus button or icon on the search page. In Ovid, terms are automatically mapped to “Subject Heading” unless users search with the box next to “Map Term to Subject Heading” empty. It helps to standardize terms and to construct precise and effective keyword or phrase searches.

Stemming

Stemming is available in ACM Digital Library, Emerald, Ovid, Ei Village, and IEEE Xplore. Using the root word of the search term as the stem basis, it retrieves items with all stem variations. ACM Digital Library and Emerald automatically stemmed the search terms. Ei Village and Ovid require the dollar sign (\$) before the search terms; and IEEE Xplore requires <stem > command in order for search terms to be stemmed.

All these retrieval features are offered by many of the database providers; however their implementation and interpretation are different.

Unique retrieval features

The retrieval features that are offered by less than five database providers and applauded by users included:

- journal browsing;
- table of content browsing;
- explode or expand search;
- lateral searching;
- density of terms;
- frequency of terms;
- reference links, persistent links, and custom links;
- searching via table of content and classification codes;
- punctuation marks, hyphenated words, and special characters (e.g. superscript, subscript, symbols);
- spell-checker;
- times an article being cited and link to the article;
- ability to browse subject authority;
- PIC variant searching (plurals, international spelling variants, and compound words);
- suggest subject headings;
- treatment types;
- automatic translation software;
- article type;
- query by example; and
- unary NOT.

Journal browsing and table of content browsing

Ovid, IEEE Xplore, Emerald and ScienceDirect provide Journal Browsing and Table of Content Browsing for journals selected by users. Users can click on the “Journals” button available on the main search page of the databases or enter the first few words of the title and select the desired journal from the displayed titles.

Explode

Ovid’s explode search feature available in the “Tool” option allows a search term to be automatically “ORed” with all of its conceptually narrower terms. Explode feature includes the term as a subject heading, heading as the focus of the article, and in combination with a subheading. This feature is best used for comprehensive retrieval.

Expand search

Gale’s expand search feature allows users to expand their search to include articles in the areas of a particular subject. Users can select “other subdivisions” links or “Other articles in this issue” to see more articles under the same subject.

Lateral search

Lateral search as offered in Ebscohost allows the selection of additional search terms from documents or records retrieved from earlier search. Searches can be narrowed through selection of more specific subject terms from the results displayed.

Density of terms

The “many” feature in IEEE Xplore allows searches of documents with a high density of a search term. The “frequency of terms” feature in Ovid allows searching documents with terms occurring at n times or more.

Links feature

SMARTLinks in Ebscohost help users in getting full-text items or articles in other Ebscohost databases or e-journals. Persistent links retrieve an article via clicking on a link embedded in a Web site. Custom links provide a link with the library’s OPAC.

Punctuation marks

Gale has a punctuation mark rules in its system. A hyphen (-) used between two words, ampersand (&) and period (.) are ignored but if searching for a word or phrase that normally contains the punctuation marks, they can be included and enclosed in double quotation marks. Apostrophes should be used when searching contractions. For possessives, the apostrophe may be used in search phrases because the search engine will return results containing the words from the query.

Unary NOT

The unary NOT feature for example for the search “a NOT b” exclude records like Boolean NOT, however without involving any condition like the term “b” being conditional on “a”. An example of a search with unary NOT is given in Figure 1.

PIC variant searching

PIC variant searching (plurals, international variants, and compound words) provided by Biblioline really help the users more than stemming and truncation. This three-in-one-feature searches for all these variants and retrieves many more relevant records for the users.

Query by example

Query by example provided by ACM Digital Library allows users to look for other documents that are similar to documents selected by users. To perform a query by example search, users need to highlight information in the document, copy and paste it into the entry field in the Web form, and click the search button. The query by example feature works best when terms relevant to

the needed documents are present in the sentences or paragraphs entered.

Others

Automatic translation software, ability to browse subject authority, suggest subject headings, times an article has been cited and spell-checker are features offered by Ebscohost that are not commonly applied in many databases, and users wish that they were incorporated in other databases too. However, users can only read about these features in the Help menu and they were unable to experiment with these features by themselves.

Users’ first impression and users’ perception after use

All 40 users expressed their first impression of the database features, which are as follows:

- simple and clear interface design;
- convenient guides are provided;
- user-friendly interface for search;
- powerful contents;
- not many screens to move;
- easy to use Boolean;
- easy to use basic search;
- advanced search for search with many options;
- proximity feature for control in searching;
- easy to get through and start searching; and
- quick downloading time.

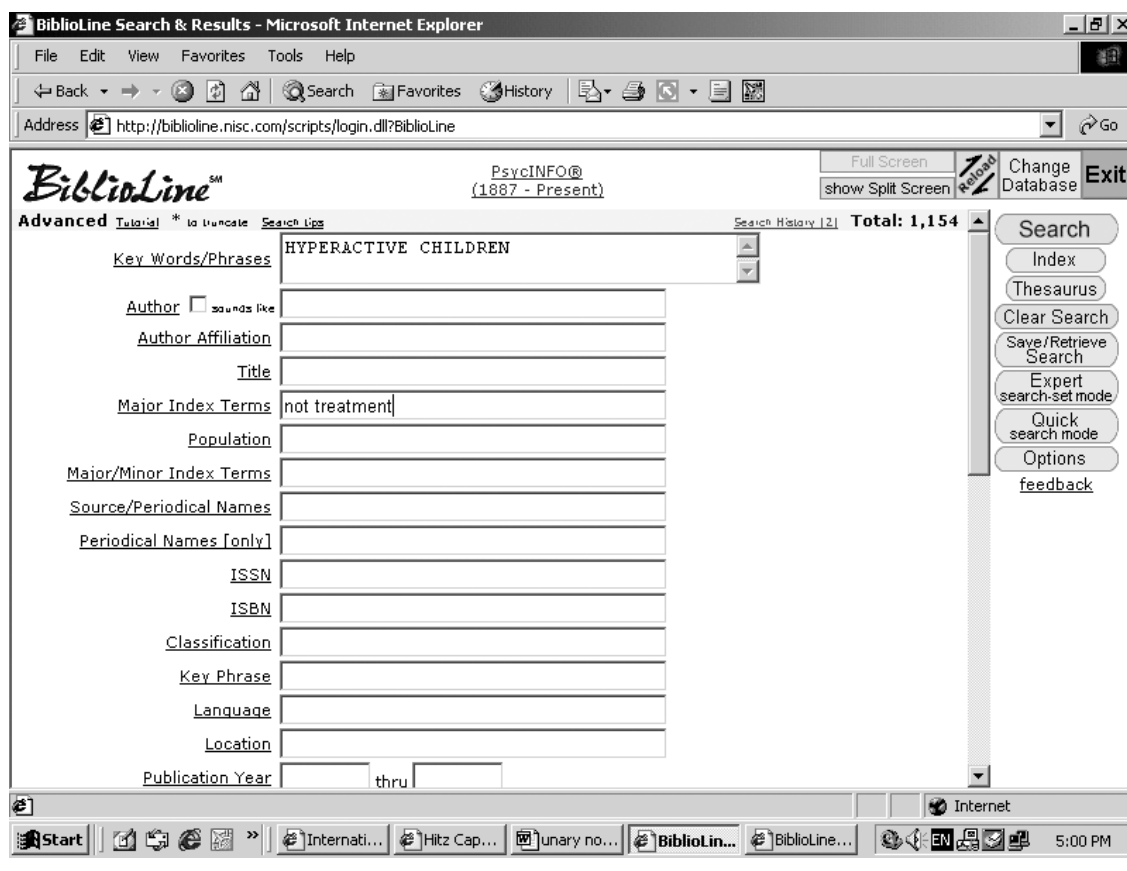
The undergraduate students were impressed with features that are general in nature, and which were search interface, Boolean operators, and basic search. The master students commented on the search interface, options of the advanced search, number of screens involved in searching and browsing, and guides. The PhD candidates commented on search interface, proximity search and control in searching, options of the advanced search, content of the databases, and guides. Basically, these different kind of users commented on different levels of retrieval features. Comments on simple and general features were made by undergraduates, while comments on specific ones were given by PhD candidates.

All of these users gave similar remarks on ease of starting the search and quick downloading time.

However, when they completed their retrieval tasks, all users regardless of different level of study gave their remarks as follows:

- interface not really user-friendly, e.g. advanced search;
- smallest change in searching affects the results;
- advanced search only benefits skilled users;
- difficulty in selecting the best terms from the Thesaurus;

Figure 1 Unary NOT is used to exclude citations with "treatment"



- should improve downloading time, e.g. pdf format; and
- screen too small e.g. ACM digital library, which requires zooming.

Users in the control group, who were also interviewed for comments before and after their searches, mentioned that they were not familiar with the databases and their retrieval features, except for Boolean operators to which they responded “okay”.

Difficulties faced by users

Difficulties faced by users in utilizing the retrieval features were related to the application of the retrieval features to searches. The implementation of the retrieval features is difficult for users since these features vary from system to system.

Users, especially the undergraduate and master students, faced difficulties in finding synonyms from the thesaurus. The thesaurus only displays broader, related, and narrower terms. Thus the users could not re-formulate their search strategies with synonyms and pseudo-synonyms.

Ebscohost listed some features in the help menu which were not found in the database; for instance, automatic translation software; ability to browse subject authority; suggest subject headings;

times an article being cited; and spell-checker. Users, especially the master students and PhD candidates, who were enthusiastic to experiment with these unique features, were completely frustrated when they discovered that the features were not there as claimed.

Users also expected to see more search examples other than the Help, Guide, or Tutorial menu and depended greatly on the search examples before starting with their own searches. They needed more help with search terms and the best application of the retrieval features to searches. Users admitted, and were even observed, to be struggling to come up with the best search terms.

There were also difficulties that were related more to configuration or subscription problems of the institution. Users had problems with downloading time, especially for documents with.pdf format, that restrain them from continuing their other searches. Another obstacle is the connection problems, particularly in LISANet and Ovid. While conducting retrieval tasks, LISANet was inaccessible due to some problems with the University’s subscription procedure; while for Ovid, users need to re-login when leaving the page idle for three minutes. Ovid displays a message that the number of

users exceeds the access limit and the user needs to re-login, without an initial warning.

Difficulties faced by users in the control group were basically related to application of Boolean operators to search terms and in one case, avoiding stopwords in her searches.

Retrieval features expected by users

Users, especially the master students and PhD candidates, expressed their wish that unique features like query by example (ACM Digital Library) should be incorporated in many other databases as well. They require the search Help Menu and Guide to give simple instructions, are not too technical and provide more examples for every command.

IEEE Xplore was rated by all 40 users as the best system in term of its retrieval features, especially the proximity search and density of terms. In addition, users felt that they had better control in searching, could search using natural language terms, and be specific in their searches. Unfortunately, IEEE Xplore offers collections related to engineering only. ACM Digital Library and EI Village were rated as second and third.

Users, especially the master students and PhD candidates, expected extensions of the proximity feature, such as those offered by ACM Digital Library and IEEE Xplore, to also be available in other databases, especially those related to Library and Information Science subject.

Other features expected by users included relevance feedback (30 of 40 users) and

assignment of weight values for search terms (27 of 40 users). These features will help users with the search terms.

Table VIII shows the features of databases other than IEEE Xplore that users expected to be maintained together with Boolean operators, phrase search and exact words, field specific search, truncation, and wild card. Most of these features were related to identification of appropriate search terms using thesaurus, topics and subjects, and indexes; search formulation; and full-text availability.

All 30 users in the control group mentioned that they need help with search formulation and search terms, and suggested that these systems should provide “suggested keywords” and “suggested search strategy” features.

Conclusion

The retrieval features that are available in many systems can now be seen as basic features to trained users. Such users expect more of the unique features that perhaps are still considered as too advanced by the database providers, which could be true with users without training and an appropriate level of search skills. This study proved that trained users do expect these “advanced” features to be available in many systems. It is now time for the database providers to consider offering features meant for advanced users, and to redefine this category of users.

Table VIII Features suggested by users to be maintained

	Database provider	Databases	Features
1.	Ovid	Medline	Medical subject heading
2.	ProQuest	ABI/Inform Global Health and Medical Complete Eric plus Text	Browse topics and subjects Thesaurus Full-text availability
3.	SilverPlatter	Education Journals Sociological Abstract EconLit	Thesaurus Thesaurus and search builder for both databases
4.	EI Village	Compendex Inspec	Controlled terms, browse indexes, and expert search for both databases
5.	Gale	Legal Trac Expanded Academic ASAP	Advanced search for both databases
6.	Biblioline	PsycArticles PsycInfo AIDSearch Child Abuse, Child Welfare and Adoption Info-ASEAN and Pacific Rim	Full-text availability and journal browsing Thesaurus for all databases
7.	Ebscohost	ERIC	Thesaurus, lateral searching, and availability of image collection
8.	ScienceDirect	ScienceDirect	Search subject areas
9.	ACM	ACM Digital Library	Stemming and full-text availability
10.	Emerald	Emerald Fulltext	Full-text availability

Extensive search examples should be given in the database.

Common features are quite standard by principle and yet their application in constructing the query varies from one search to another, and from one system to another. Unique features are also useful for other databases since users do not limit their searches to one database only. Expected features, like relevance feedback and term weighting, are already implemented in systems developed internally at institutions. It is best that these features are also incorporated in these databases by their providers.

Users' difficulties in using these databases are still related to searching. Therefore, examples of features application with the subject's term are crucially needed. Technical difficulties like downloading time and number of users exceeding access limits are related to the configuration and subscription as undertaken by the institution offering access to the databases, and thus a change in policy may be needed. In any study, users' first impressions must be separated from users' perceptions after use, as these are very different.

References

- Brajnik, G., Mizzaro, S., Tasso, C. and Venuti, F. (2002), "Strategic help in user interfaces for information retrieval", *Journal of the American Society for Information Science*, Vol. 53 No. 5, pp. 343-58.
- Diercks, T. (2003), "Database evaluation report: Expanded Academic ASAP vs Academic Search Premier", available at: http://libweb/hawaii.edu/uhmlib/news/surveys/eai_asap.htm.
- Eastman, C.M. and Jansen, B.J. (2003), "Coverage, relevance, and ranking: the impact of query operators on Web search engine results", *ACM Transactions on Information Systems*, Vol. 21 No. 4, pp. 383-411.
- Electronic Collections Committee (1993), "Evaluation of UMI's Proquest Education Complete", available at: www.fcla.edu/sulc/ecc/97_99/umieval.html
- Heidig, L. and Prior, B. (1997), "Aggregator interface evaluation report", available at: www.englilb.cornell.edu/cul/a2i/reports/interface_eval.html
- Kline, V.L. (2002), "Missing links: the quest for better search tools", *Online Information Review*, Vol. 26 No. 4, pp. 252-5.
- Ma, W. (2002), "A database selection expert system based on reference librarian's database search strategy: a usability and empirical evaluation", *Journal of the American Society for Information Science*, Vol. 53 No. 7, pp. 567-80.
- Robins, D. (2000), "Shifts of focus on various aspects of user information problems during interactive information retrieval", *Journal of the American Society for Information Science*, Vol. 51 No. 10, pp. 913-28.
- Schaffer, T. (2001), "Databases and political science research", *Online Information Review*, Vol. 25 No. 1, pp. 47-53.
- Shiri, A.A., Revie, C. and Chowdhury, G. (2002), "Thesaurus-enhanced search interfaces", *Journal of Information Science*, Vol. 28 No. 2, pp. 111-22.
- Shneiderman, B., Byrd, D. and Croft, W.B. (1998), "Sorting out searching: a user-interface framework for text-searches", *Communication of the ACM*, Vol. 41 No. 4, pp. 95-8.
- Sutcliffe, A.G., Ennis, M. and Hu, J. (2000), "Evaluating the effectiveness of visual user interfaces for information retrieval", *Journal of Human-Computer Studies*, Vol. 53 No. 5, pp. 741-63.
- Tenopir, C. and Hover, K. (1993), "When is the same database not the same? Database differences among systems", *Online*, Vol. 17 No. 2, pp. 20-8.
- Wolfram, D. and Xie, H. (2002), "Traditional IR for Web users: a context for general audience digital libraries", *Information Processing and Management*, Vol. 38 No. 5, pp. 627-48.
- Xie, H. (2003), "Supporting ease-of-use and user control: desired features and structure of Web-based online IR systems", *Information Processing and Management*, Vol. 39 No. 6, pp. 899-922.
- Xie, H. and Cool, C. (2000), "Ease of use versus user control: an evaluation of Web and non-Web interfaces of online databases", *Online Information Review*, Vol. 24 No. 2, pp. 102-15.