

City University of New York (CUNY)

CUNY Academic Works

Publications and Research

York College

2010

Standardization of Keyword Search Mode

Di Su

York College

[How does access to this work benefit you? Let us know!](#)

More information about this work at: https://academicworks.cuny.edu/yc_pubs/2

Discover additional works at: <https://academicworks.cuny.edu>

This work is made publicly available by the City University of New York (CUNY).

Contact: AcademicWorks@cuny.edu

Standardization of Keyword Search Mode

Abstract

In spite of its popularity, Keyword Search mode has not been standardized. While information professionals are quick to adapt to various presentations of Keyword Search mode, novice end-users often find Keyword Search confusing. This article compares Keyword Search mode in some major reference databases and calls for standardization.

KEYWORDS. Bibliographic database, database interface, online public access catalog, OPAC, reference database, search method, search mode, search syntax, standardization

Introduction

Keyword Search is one of the search modes in online catalogs and bibliographic databases. Unless one has a specific item at hand, such as a title, an author, or other traditional access points, the searcher faces options of using either Subject Search or Keyword Search. Subject Search utilizes controlled vocabulary, which consists of searchable terms selected by an authoritative agency, e.g. the Library of Congress Subject Headings. In comparison, Keyword Search offers more flexibility in formulating search statement, and is less strict in search terms. Because of its convenience, Keyword Search is set as default search mode in most databases. Although the creation of Z39.50, an international standard for communication between two computer systems, has popularized uniformity in database interface to some extent, Keyword Search mode has not been formally standardized at a more meaningful level. While information

professionals and savvy researchers are quick to adapt to various presentations of Keyword Search mode, novice end-users often find that Keyword Search demands too much time and knowledge in navigating among various reference databases.

This article focuses on structural settings of Keyword Search mode in some major reference databases, comparing functionalities, illustrating differences, and recommending items for standardization. It should be noted that this article is not confined to Basic Search option only, as some features in Advanced Search option are also relevant, such as Boolean Logic and truncation. It is hoped that this article will draw much-needed attention to the standardization of Keyword Search mode.

Background

When a library user looks for a book, or a journal article, or literature on a subject, he or she usually utilizes information retrieval systems. In the old days, card catalogs and print indexes were standard information retrieval systems. They provided so-called access points that would lead the user to the desired item(s). Traditional access points include the title, author, and subject headings (or descriptors in some indexes, e.g. in ERIC's *Current Index to Journals in Education*). While some print indexes are still in use, card catalogs are rarely seen in today's libraries.¹ They have been almost entirely replaced by online public access catalogs (OPAC). For example, the content of the Library of Congress's *National Union Catalog* (NUC), the monumental compilation of card catalogs in American and Canadian libraries, is now available in Online Computer Library Center's (OCLC) WorldCat, which contains a collection of records in 71,000 libraries of 112 countries (as of April 20, 2009).² One can find out "who has what in

where” at a computer terminal through the online WorldCat instead of flipping through the heavy volumes of NUC. Meanwhile, more and more indexes and abstracts also become available electronically. For example, Medline covers the content of hard copies of *Index Medicus*, and PsycINFO offers online version of *Psychological Abstracts*. Finally, the era of full-text bibliographic databases has arrived.

What is new in this online environment? One of the most evident advantages of computerized catalogs and bibliographic databases is the availability of keyword as an additional access point. Keyword(s) is a significant word (or phrase) not only in the title, author, or subject headings (or descriptors), but also in the content notes, abstract, or text of a record, in an online catalog or bibliographic database. A keyword search covers a similar but broader range than a subject search. Keyword(s) can be used as a search term in a free-text search (using natural language words and phrases), or in a full-text search (scanning the entire document), to retrieve all the records that contain it. Keyword Search opens up a new, powerful way of information seeking and retrieving. It becomes an essential search method that is particularly useful for the end-users, who have little information on access points by traditional means, but a broad idea on a research topic.

The Need

As a tradeoff, the convenience of Keyword Search comes with drawbacks. Among many issues, the lack of precision and relevancy seems to have received most complaints, (e.g., a Keyword Search may produce irrelevant records, known as false drops, whenever the search term happens to have more than one meaning.) Standardization of Keyword Search mode, on the other hand, has not received much-

deserved attention by its own right. A Keyword Search may unexpectedly fail due to the lack of standardized methods. For example, the searcher may inadvertently break hidden rules in some reference databases, such as the use of unsupported natural language or forbidden stopwords. The failed search could be so discouraging that the searcher may turn to other sources that are more user-friendly. (It seems to explain why some end-users begin their research by choosing Google, for its simplicity, over academic reference databases.) Standardization of Keyword Search mode in various reference databases would simplify the search method and save end-user's time.

In spite of the popularity that Keyword Search enjoys, there has been very little literature on calling for standardization of Keyword Search mode. Carol Tenopir predicted in 2002, "standardization isn't likely to happen anytime soon."³ She was referring to the syntax for proximity operators in particular, but her comment could be applied to Keyword Search mode in general as well. The current situation seems to validate Tenopir's prediction. In the same article, she vividly described a "near-panic" student who shouted for help, "Factiva, LexisNexis, Westlaw, Dialog, ProQuest, CSA...they are all running together! I just can't keep track of them!"⁴ Tenopir did not specify what kind of problem it was, but it seems probable that a search syntax or method that obtained search results from one database, did not work in the others. The problem could be minimized if the standardization of database search methods becomes a reality.

Another drawback caused by the lack of standardization is the inconsistency. With an identical search statement and the same content source, simultaneous searches on different platforms may retrieve different results. An example is OCLC's WorldCat. The same WorldCat search performed on WorldCat.org (public version, available at

<<http://www.worldcat.org/>>) and WorldCat/FirstSearch (subscription required) produces different results. (See Table 1: WorldCat on Different Platforms). The main reason, OCLC explains, is that “the default order of results differs between the WorldCat.org interface and the FirstSearch interface.”⁵ OCLC further explains, “Results are displayed according to relevance in WorldCat.org, so the items most closely related to a user's search terms will appear at the top of the results. In the FirstSearch view of WorldCat, results are displayed according to the number of holdings for each item. Therefore, items that are owned by the most WorldCat libraries will appear at the top of the results.”⁶ In addition, the designs of search methods could affect the outcomes, e.g., while Boolean Logic is accepted in WorldCat/FirstSearch, it is not supported in WorldCat.org. (For more database examples of this sort, see Table 2: Same Database Source on Different Platforms.)

Variations among reference databases are expected by librarians and information professionals due to the job nature. For novice end-users, however, the variations may be burdensome, to say the least. The user's frustration often prompts librarians to wonder, “if there is a simple way to keep online systems straight in order to help users and conduct efficient searches.”⁷ Again, standardization can be the answer.

Comparison

To compare the variations in Keyword Search mode among some major reference databases, sample searches were performed for this article. The following methods were applied to assure the consistency:

1. As Web page presentations may vary in different Web browsers, the sample

searches were performed in Internet Explorer only.

2. It was on the same day that all the sample searches were performed, and all the relevant Web addresses were checked.

3. Identical search statements were used in all the databases listed below:

WorldCat / OCLC FirstSearch

WorldCat / OCLC public version

ERIC / EBSCO

ERIC / CSC public version

Medline / EBSCO

Medline / NLM public version

Academic Search Premier / EBSCO

Cambridge Journals Online

Health Reference Center Academic / InfoTrac/Gale

JSTOR

Lexis-Nexis

Project MUSE

ProQuest Historical Newspapers: *The New York Times* (1851-2004)

Readers Guide Full Text Mega Edition / H.W.Wilson

SAGE Journals

ScienceDirect

Wiley InterScience

It should be noted that the same search statements in the same databases may produce different outcomes day to day, due to: a) rapid change in technology; b) database

growth and improvement; c) unforeseen events in digital publishing business, e.g. merger of companies.

Discussion

Standardization is “the process of establishing uniform procedures and standards in a specific field of endeavor, usually to facilitate exchange and cooperation and to assure quality and enhance productivity. In librarianship, standards are established by professional associations, accrediting bodies, and government agencies.”⁸

To standardize Keyword Search mode, an agency with given authority needs to collaborate with library communities and database producers, gathering and synthesizing relevant information, evaluating proposals and suggestions, and establishing and recommending the standards. Successful examples of standardization include the works of the International Federation of Library Associations and Institutions (IFLA), which produced a set of cataloging rules based on the International Standard Bibliographic Description (ISBD),⁹ and National Information Standards Organization (NISO), which set standard for Z39.50.¹⁰

Keyword Search is set as default search mode by most reference databases for practical reasons. However, the lack of standardization in Keyword Search mode can be seen in many facets. With end-users in mind, issues for consideration of standardization may include, but not limited to, the name of Keyword Search mode, Natural Language Search, Boolean Logic, truncation and wildcard, Help index, and Z39.50. (For current status, see Table 3: Database Comparison.) The goal is to make the Keyword Search mode more intuitive and user-friendly through standardization. The following is the

discussion on the issues.

Name of Keyword Search mode

There are various names for Keyword Search mode, such as Basic Search, Quick Search, Easy Search, Article Search, and so forth. Let us designate a uniform name for Keyword Search as long as it is distinguishable from the Title Search, Author Search, and Subject Search. The name should be universally understandable, not only to librarians and information professionals but also to end-users, that it is a true free-text or full-text search. Preference goes to “Keyword Search” for its clarity.

Natural Language Search

In comparison with command-driven Boolean Search, Natural Language Search employs relevance ranking capabilities and intelligent text processing search engines. These “Post-Boolean” search engines, as Barbara Quint called them, “use a complex series of algorithms to analyze statistical counts of terms (the number of terms in each document, frequency of terms in document compared to frequency of terms in database, etc.).”¹¹ Early products of Natural Language Processing (NLP) systems included CQ’s Washington Alert, Dow Jones News/Retrieval’s DowQuest (both launched in 1989), and Westlaw’s Westlaw Is Natural (WIN) (in 1992). The NLP systems “are based on the assumption that our standard command-driven online systems coupled with Boolean logic searching are not only difficult to learn, but may sometimes miss relevant documents.”¹² In her article in 1996, Susan Feldman described the concept with end-users in mind: “Ideally, an information retrieval system should be an answer machine. It should

interpret the questions we ask in plain English and return not only what we asked for, but what we meant to ask for.”¹³

Natural Language Search is now available in almost all Web search engines, but it is not a standard feature yet in OPACs and bibliographic databases. Evidently, the Internet is changing the information world from an environment, which used to be controlled exclusively by information professionals, to one that is more end-user-centered. End-users are so accustomed to Web search methods that they may search in the same way when using reference databases, where Natural Language Search statements may or may not be accepted. “Natural language is easier for end-users to use and it can outperform Boolean,”¹⁴ commented Nicholas Tomaiuolo and Joan Packer, when they compared the two search modes, “while Boolean searches are precise, natural language searches are comprehensive.” They concluded, “Searchers should appreciate the additional power and retrieval of natural language searching.”¹⁵ The jargon-free and stopwords-tolerant Natural Language Search capability could be a relief for end-users. Several bibliographic database vendors within the scope of this study, such as EBSCO, Lexis-Nexis, and Wilson, have recognized the importance of Natural Language Search, and implemented the feature in their products already. Preference goes to Natural Language Search support.

Boolean Logic

The majority of tested databases use a pull-down menu for Boolean Logic operators in Advanced Search option. When typing is necessary, some databases require that operators, “AND”, “OR”, and “NOT”, be capitalized, while others do not. Since the

words “and”, “or”, and “not” may also appear in Natural Language Search statements with non-Boolean functions, the capitalization of Boolean Logic operators seems necessary. In some databases, the form of “AND NOT”, which simply means “NOT”, is a rather awkward usage of Boolean Logic operator. It is unnecessary to make an already jargon-bearing concept more confusing to end-users. Preferences go to operators “AND”, “OR”, and “NOT” with capitalization.

Truncation and wildcard (a.k.a. wild card)

We see the asterisk “*” as a popular symbol for truncation and the question mark “?” as a common choice for wildcard. There are also various symbols in use, such as the number sign “#” for wildcard in WorldCat/FirstSearch, and the exclamation mark “!” for truncation in Lexis-Nexis. Other variations for truncation and wildcard in the databases beyond this study include the dollar sign “\$”, the colon mark “:”, and the plus sign “+”. Currently, as the truncation and wildcard symbols are not standardized, users are advised to read the Help section, especially in an unfamiliar interface, to make sure: a) if truncation and wildcard are available; b) if they are, what symbols are designated as truncation and wildcard respectively. Standardizing the symbols would make things easier to remember. Among the variations, the dollar sign “\$” seems to bear the least sense since it is generally used for monetary matters. The plus sign “+” is commonly associated with mathematical functions. The question mark “?” is a *de facto* popular symbol in practice, except its location on keyboard is a distance away from the group of common symbols.

The more suitable ones seem to be the asterisk “*” and the number sign “#”. The

asterisk “*” is already used in most operating systems and application programs in computer science as “wildcard”. It can be used in a filename to identify multiple files and directories, e.g., use *cat*.doc* to retrieve all the *doc* files with filenames beginning with “*cat*”. Its function, although called “wildcard” in computer science, is similar to the concept of truncation in library science. The number sign “#” seems more straightforward when it is designated as wildcard to replace a “number” of letter(s). In short, the concepts of truncation and wildcard are already complicated, let us make it as simple as possible for end-users. Preferences go to the asterisk “*” for truncation and the number sign “#” for wildcard.

Help index

Online Help section is often consulted by searchers when things are uncertain. There are two kinds of Help index, searchable and browseable. Some databases offer searchable Help index, which is particularly handy, while others only provide non-searchable Help index for browsing, which is often time-consuming to use. Preference goes to searchable Help index.

Z39.50

Z39.50 is a client-server protocol that allows a user in one computer system to search and retrieve information from others (also Z39.50 implemented), and to receive results in the format of the local computer system. To some extent, Z39.50 has created a framework for standardization of Keyword Search interface, since some online catalogs and most bibliographic databases already support Z39.50 standard.¹⁶ Preference goes to

Z39.50 compatibility.

Conclusion

Keyword Search in online catalogs and bibliographic databases is a crucial search mode for end-users. With too many variations, its current search method is inefficient and confusing due to the lack of standardization. There may be numerous reasons that database producers define and design Keyword Search syntax and methods in various ways. For end-users, however, we can see instant benefits from standardizing Keyword Search mode. The standardized Keyword Search mode would lead to a more user-friendly search environment, affect end-user's search experience in a positive way, hence encourage and increase the use of academic reference databases.

Table 1: WorldCat on Different Platforms

Platform	OCLC http://www.worldcat.org/	OCLC FirstSearch
Search Statement A: <i>air pollution in shanghai</i>		
Number of titles retrieved	73	21
Top 5 titles	<ol style="list-style-type: none"> 1. Assessment of health benefits from controlling air pollution in Shanghai, China. 2. Low-carbon energy policy and ambient air pollution in Shanghai, China: a health-based economic assessment. 3. Quantifying the human health benefits of curbing air pollution in Shanghai. 4. The association of daily diabetes mortality and outdoor air pollution in Shanghai, China. 5. Assessment and management of urban air pollution: a new project in Shanghai. 	<ol style="list-style-type: none"> 1. China's dilemma: economic growth, the environment and climate change. 2. The law of energy for sustainable development. 3. Third International Conference on Atmospheric Sciences and Application to Air Quality: Shanghai, P.R.C., 15-19 October 1990. 4. Urban air pollution in megacities of the world. 5. The Cost of Pollution in China economic estimates of physical damage.
Search Statement B: <i>(auto* OR car?) AND (air pollution) AND shanghai</i>		
Number of titles retrieved	0	6
Top 5 titles	[N/A]	<ol style="list-style-type: none"> 1. China's dilemma: economic growth, the environment and climate change. 2. Air pollution from mobile sources in five Asian megacities Levels of air pollution and management frameworks in Bangkok, Delhi, Jakarta, Seoul and Shanghai. 3. Assessment of health benefits from controlling air pollution in Shanghai, China. 4. Planning for sustainable urban development: cities and natural resource systems in developing countries. 5. Foreign direct investment as a vehicle for deploying cleaner technologies: technology transfer and the big three automakers in China.

Table 2: Same Database Source on Different Platforms

Database name	Platform	Keyword Search mode name	Natural language support	Boolean Logic availability	Truncation symbol	Wildcard symbol	Searchable Help index	Z39.50 compatibility
<i>WorldCat</i>	OCLC http://www.worldcat.org	Search everything	No	No	? or *	#	No	No
<i>WorldCat</i>	OCLC FirstSearch	Basic Search	No	Yes	*	# (represents a single character) ? (alone or with a number, represents from zero to nine additional characters)	Yes	Yes
<i>ERIC</i>	CSC http://www.eric.ed.gov	Basic Search	No	Yes	*	*	Yes	No (uses OpenURL Z39.88)
<i>ERIC</i>	EBSCO	Basic Search	Yes (termed “SmartText Searching”)	Yes	*	?	No	Yes
<i>Medline</i>	NLM http://www.ncbi.nlm.nih.gov/sites/entrez?db=PubMed	Search	Yes	Yes (must use uppercase characters for operators)	*	N/A	Yes	No
<i>Medline</i>	EBSCO	Basic Search	Yes (termed “SmartText Searching”)	Yes	*	?	No	Yes

Table 3: Database Comparison

Database name	Keyword Search mode name	Natural language support	Boolean Logic operators	Truncation symbol	Wildcard symbol	Searchable Help index	Z39.50 compatibility
<i>Academic Search Premier (EBSCO)</i>	Basic Search	Yes (termed "SmartText Searching")	AND, OR, NOT. (type in or use pull-down menu)	*	?	No	Yes
<i>Cambridge Journals Online</i>	Quick Search	No	AND, OR, AND NOT. (use pull-down menu)	* (termed "Wild Card")	*	Yes	No
<i>Health Reference Center Academic (InfoTrac/Gale)</i>	Basic Search	No	AND, OR, NOT. (use pull-down menu)	* (termed "Wildcard")	? (represents one character) ! (represents one or no character)	No	Yes
<i>JSTOR</i>	Basic Search	No	AND, OR, NOT. (use pull-down menu)	* (termed "Wildcards")	?	No	No (uses XML gateway)
<i>Lexis-Nexis</i>	Easy Search	Yes	AND, OR, AND NOT. (type in in Power Search)	!	*	No	No

Standardization of Keyword Search Mode

<i>Project MUSE</i>	Article Search	No	AND, OR, NOT. (type in or use pull-down menu)	*	No	No	Yes
<i>ProQuest Historical Newspapers: The New York Times (1851-2004)</i>	Basic Search	No	AND, OR, AND NOT. (use pull-down menu)	*	?	Yes	Yes
<i>Readers Guide Full Text Mega Edition (Wilson)</i>	Basic Search	Yes (termed “All-Smart Search”)	AND, OR, NOT. (use pull-down menu)	*	?	Yes	Yes
<i>SAGE Journals</i>	Quick Search	No	AND, OR, NOT. (use pull-down menu)	*	?	No	No
<i>ScienceDirect</i>	Quick Search	No	AND, OR, AND NOT. (type in Expert Search or use pull-down menu)	* (termed “Wildcard”)	?	Yes	No (uses XML gateway)
<i>Wiley InterScience</i>	Search	No	AND, OR, NOT. (use pull-down menu)	* (termed “wild character asterisks”)	?	No	No

References

- Feldman, Susan. "Testing Natural Language." *Online* 20, no.6 (November/December 1996): 71-79.
- Hock, Ran. "How to Do Field Searching in Web Search Engines." *Online* 22, no.3 (May/June 1998): 18-22.
- . "Revisiting Web Search Engines: Features and Commands." *Online* 25, no.5 (September/October 2001): 18-24.
- Ojala, Marydee and Ernest Perez. "Web Search Engines: Search Syntax and Features." *Online* 26, no.5 (September/October 2002): 28-31.
- Quint, Barbara. "The Artifices of Natural Language Searching." *Wilson Library Bulletin* 69 (December 1994): 60-61.
- Tenopir, Carol. "Sorting Through Online Systems." *Library Journal* (May 1, 2002): 32-34.
- Tenopir, Carol, Gayle Baker, and William Robinson. "The Information Playground." *Library Journal* (May 15, 2004): 38-42.
- Tenopir, Carol and Pamela Cahn. "Target & Freestyle: DIALOG and Mead Join the Relevance Ranks." *Online* 18, no.3 (May 1994): 31-47.
- Tomaiuolo, Nicholas G. and Joan Packer. "Maximizing Relevant Retrieval: Keyword and Natural Language Searching." *Online* 22, no.6 (November/December 1998): 57-60.

Endnotes

-
- ¹ Samples of card catalog can be viewed online at <<http://gustavus.edu/academics/library/libraryhistory/collection05.html>>. (Courtesy of the Gustavus Adolphus College, Folke Bernadotte Memorial Library, Saint Peter, Minnesota.) [Accessed April 20, 2009].
- ² “WorldCat facts and statistics.” Available: <<http://www.oclc.org/worldcat/statistics/default.htm>>. [Accessed April 20, 2009].
- ³ Carol Tenopir, “Sorting Through Online Systems,” *Library Journal* (May 1, 2002): 32.
- ⁴ Ibid.
- ⁵ “WorldCat.org frequently asked questions.” Available: <<http://www.oclc.org/us/en/support/questions/worldcatorg/default.htm>>. [Accessed April 20, 2009].
- ⁶ Ibid.
- ⁷ Tenopir, “Sorting Through,” 32.
- ⁸ Joan M. Reitz, *ODLIS — Online Dictionary for Library and Information Science*. Available: <<http://lu.com/odlis/>>. [Accessed April 20, 2009].
- ⁹ The International Standard Bibliographic Description (ISBD) is intended to serve as a principal standard to promote universal bibliographic control. The ISBD's main goal is to offer consistency when sharing bibliographic information. Current edition of ISBD is available online at <<http://www.ifla.org/VII/s13/pubs/cat-isbd.htm>>. [Accessed April 20, 2009].

¹⁰ For more information, visit the Web site of the Z39.50 standard's maintenance agency, Network Development & MARC Standards Office of the Library of Congress, at <http://www.loc.gov/z3950/agency/>. [Accessed April 20, 2009].

¹¹ Barbara Quint, "The Artifices of Natural Language Searching," *Wilson Library Bulletin* 69 (December 1994): 61.

¹² Carol Tenopir and Pamela Cahn, "Target & Freestyle: DIALOG and Mead Join the Relevance Ranks," *Online* 18, no.3 (May 1994): 31.

¹³ Susan Feldman, "Testing Natural Language," *Online* 20, no.6 (November/December 1996): 71.

¹⁴ Nicholas G. Tomaiuolo and Joan Packer, "Maximizing Relevant Retrieval: Keyword and Natural Language Searching," *Online* 22, no.6 (November/December 1998): 58.

¹⁵ *Ibid.*, 60.

¹⁶ It is worth noting that instead of implementing Z39.50 protocol, ScienceDirect and JSTOR have deployed an XML gateway, which, according to ScienceDirect, "fully unlocks its native search capability." ("ScienceDirect Content, Linking, and Software FAQs" in online Help.) ERIC/CSC (Computer Sciences Corporation), on the other hand, implemented support for OpenURL (Z39.88 standard, for which OCLC is the maintenance agency) in the form of its "Find in a Library" feature.