THE DISAPPEARANCE OF CONTROLLED VOCABULARY IN BIBLIOGRAPHIC INSTRUCTION: *IN MEMORIAM*?

STEPHANIE WIEGAND

INTRODUCTION

In the 4,500 years between the first clay tablet and the first ebook (O'Gormann, 2012, p. 2), much has changed. Human written knowledge started with a very small amount of information stored in very large physical containers and now extremely large amounts of information are stored in very small storage systems. The 17th century saw the publication of scientific journals, the 18th century witnessed the creation of the first indices of scientific and law journals, and the 20th century observed the advent of electronic and online indices and content (Wiegand, 1994, 574, 269, 152). It is estimated that humans now produce one to two exabytes of information each year (Lyman, 2000, para. 1). As our knowledge expands, we pioneer new methods of communicating information and delving into the wealth of information that exists to find the one piece that we need at a specific point in time. Access to information was the primary obstacle to users throughout much of history. Today, at times, access is less an issue than is discoverability, simply due to the sheer volume of what searchers must sort through to find the specific information needed.
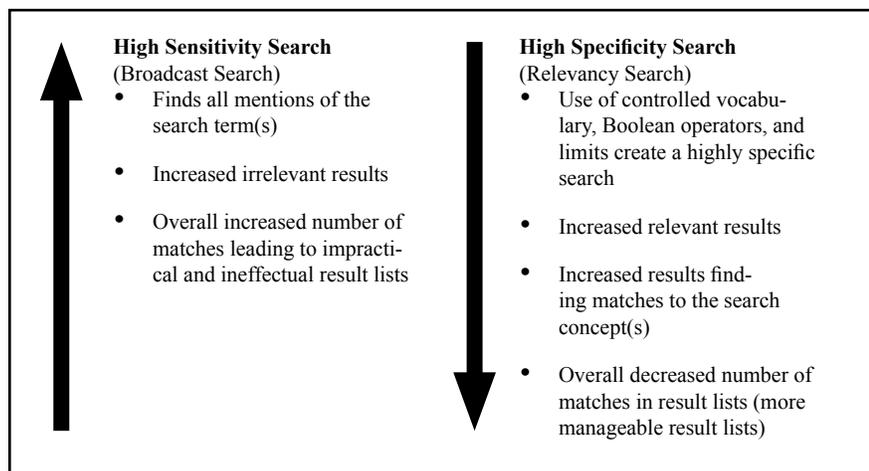
Most print indexes were simple to use, though the tiny print and abbreviations stymied many irresolute researchers. Print indexes offered high search specificity, but low search sensitivity, which meant many relevant sources of information were missed in searches of the literature. Early online searches in databases such as Dialog, BRS, MEDLARS, and RLIN were fee-based, and required mediated searching where professionals input the searches for patrons as any mistakes in search queries could prove expensive (Coffmann, 2012, Intermediated Searching, para. 1). As late as 1993, librarians still believed that mediated searching was the future (Tyler & Brahmi, 1993). With the arrival of graphic user interfaces (GUI) and the marked progression in infrastructure and content on the Internet, the days of mediated searching declined rapidly. End users still often needed instruction in how to input search terms and use logic connectors, but costs of searching dramatically decreased to the point where it became more cost effective for end users to complete their own searches.

Online databases of source records or full text sources offer far greater search sensitivity than did their print counterparts, but many search components now allow for decreased specificity in searching (see Figure 1). The loss of search specificity is often due to databases that either lack controlled vocabulary schema or bury the controlled vocabulary tools in 'Help' links. In databases, controlled vocabulary is often referred to as subject headings, subject terms, descriptors, MeSH, and other terms. The use of controlled vocabulary allows the end user to create searches that retrieve highly relevant results by identifying only results that are significantly related to the topic, while also identifying results no matter what terms used by the author of the work (see Table 1). While this may not be readily apparent to the end user, authors use a variety of terminology to indicate the same phenomena, which can cause confusion when conducting database searches. If a researcher is endeavoring to identify all relevant information concerning sudden infant death syndrome, the researcher must either search for one subject heading or all of the following terms: sudden infant death, sudden infant death syndrome, sudden baby death, SIDS, cot death, and crib death. Authors use different terminology to identify the same phenomena for three

Wiegand (Health Sciences Reference Librarian & Associate Professor)
University of Northern Colorado [Greeley, CO]

Figure 1: High Sensitivity Searches versus High Specificity Searches

High Sensitivity Search
(Broadcast Search)
- Finds all mentions of the search term(s)
- Increased irrelevant results
- Overall increased number of matches leading to impractical and ineffectual result lists

High Specificity Search
(Relevancy Search)
- Use of controlled vocabulary, Boolean operators, and limits create a highly specific search
- Increased relevant results
- Increased results finding matches to the search concept(s)
- Overall decreased number of matches in result lists (more manageable result lists)

primary reasons: differences in age or generation, differences in regional or national localities, and differences in specialized knowledge. For instance, a person born in the 1920s might refer to soft drinks as fountain drinks, while someone from Alabama may refer to all soft drinks as cokes, and a person who works in the industry may refer to soft drinks as carbonated beverages.

Table 1: Keyword versus Controlled Vocabulary Searches

Type of Search & Terms	Number of Results
Keyword search for nose bleed	5
Keyword search for nosebleed	21
Keyword search for epistaxis	612
Subject Heading search for epistaxis	509

Note. Search performed in CINAHL nursing and allied health database on August 1, 2012.

The other distinct advantage that controlled vocabulary has over keyword searching is that end users are often able to identify and learn the vocabulary of the profession.

Librarians also use Boolean operators to create highly specific searches. Today many library databases employ algorithms that automatically add logic (the Boolean operator AND) between each word in search string. This automatic insertion of logic mimics Google, but often with less success and without the knowledge of the end user. So while the end user may believe she is searching with natural language or a specific phrase, many library databases now rewrite the search string to include the most common connective logic. More disturbingly, if this type of search string retrieves no matching results, some databases now automatically convert the connective logic from conjunction to disjunction, retrieving results that lack any meaningful connection to what the searcher needs.

The professional standards of academic librarianship require the teaching of Boolean logic and the measurement of student comprehension and practical performance of creating search strategies that include Boolean operators (Association of College & Research Libraries, 2012, Standard Two). The commonly used textbooks for introduction to research courses (generally taught by professional librarians) all include sections explaining the use of Boolean logic in building searches (Quaratiello, 2011; Hock, 2010; Hacker & Fister, 2010; Palmquist, 2006; Bolner & Poirier, 2007; Ercegovac, 2008; Badke, 2011; & George, 2008); however, over the last two years major library database vendors redesigned search interfaces to perform more like Google – including eliminating the necessity of understanding Boolean logic and of using Boolean operators to create search strings. Examples of major database vendors that automatically employ Boolean logic (as Google does) include all ProQuest databases, all EbscoHOST databases, all JSTOR content databases, and a host of smaller database vendors. It appears that standards and commonly accepted practices of academic librarianship may be out of step with the reality of what students need to know to proficiently search library databases. Oddly, professional communication channels – both formal and informal – remain quiet on this topic. Anecdotally, through many conversations with academic librarians, it appears that many of us now choose not to spend time teaching Boolean logic and rather devote that time to other concepts that students still need to master. The following research was undertaken to determine if professional librarians at academic institutions are still teaching Boolean logic and Boolean operators in library research credit-bearing courses and one-shot library sessions. This survey is a precursor to research investigating the teaching practices of professional librarians at academic institutions concerning controlled vocabulary and Boolean operators in library research credit-bearing courses and one-shot library sessions.

METHODOLOGY

An online survey was created with 14 questions. The survey included two demographic questions, two questions to ascertain the type of classes the respondents taught, eight questions concerning teaching of Boolean logic and Boolean operators, and two open-ended questions regarding the professional beliefs of the respondents. Email addresses were collected for 116 professional academic librarians. These librarians all currently work at four-year institutions of higher education offering both undergraduate- and graduate-level degrees. All institutions are regionally proximate to Greeley, Colorado. The online survey was open for two weeks to collect responses.

RESULTS

Only 116 professional academic librarians were contacted and asked to complete the survey. Out of the 116 professional academic librarians contacted, 41 completed the survey for a response rate of 35%. Over half of the respondents (56%) received their ALA-accredited master's degree or advanced subject-specific degree between 2000 and 2010; 23% between 1990 and 1999; 8% between 1980 and 1989; and 8% between 1970 and 1979. Although some of the newest graduates experienced some lag time between graduation and their first professional position in an academic library, most of the respondents had worked in professional positions in academic libraries since earning their degrees. The range of years of experience ranged from under one year to 34 years; the average years of experience were 11. Only 12% of respondents (5 librarians) reported teaching a library research credit-bearing course within in the last year, while 90% of respondents (37) reported teaching a one-shot library session (not including library tours) within the last year.

Of the respondents who taught library research credit-bearing courses within the last year, three (60%) reported teaching the concept of Boolean logic frequently or always, while one (20%) teach Boolean logic infrequently and one (20%) never teaches Boolean logic to students in the course. However, four of the five respondents who taught credit-bearing courses always teach the logic connective AND; one respondent teaches AND infrequently. Three respondents teach the logic connective OR always, and two never teach OR. Three of the five respondents teach negation (the Boolean operator NOT) frequently or always, while two never teach negation. See Table 2 for percentage breakdowns of the results.

Table 2: Frequency of Boolean logic & operators taught in library research credit-bearing courses

|	Concept	AND	OR	NOT
Never	20%	0%	40%	40%
Infrequently	20%	20%	0%	0%
Sometimes	0%	0%	0%	0%
Frequently	20%	0%	0%	20%
Always	40%	80%	60%	40%

Of the 37 respondents that taught one-shot library sessions in the last year, the majority are teaching Boolean logic, conjunction, and disjunction frequently or always (51%, 62%, and 54%, respectively). The reverse is true of respondents teaching the Boolean operator NOT, as the majority of respondents teach negation infrequently or never (57%). See Table 3 for the complete percentage breakdown of the results.

Table 3: Frequency of Boolean logic & operators taught in one-shot library sessions (not including tours)

	Concept	AND	OR	NOT
Never	17%	11%	14%	20%
Infrequently	11%	6%	11%	37%
Sometimes	20%	20%	20%	17%
Frequently	31%	31%	37%	14%
Always	20%	31%	17%	11%

Interestingly, while the majority of respondents teaching one-shot sessions do teach conjunction, disjunction, and the concept of Boolean logic frequently or always, it is generally a narrow majority (see Figure 2). It appears that half of the respondents teach the concept of Boolean logic and disjunction, while a greater number teach conjunction.

Figure 2: Comparison of the Frequency Boolean Logic & Operators Are Taught in One-Shot Library Sessions (Not Including Tours)
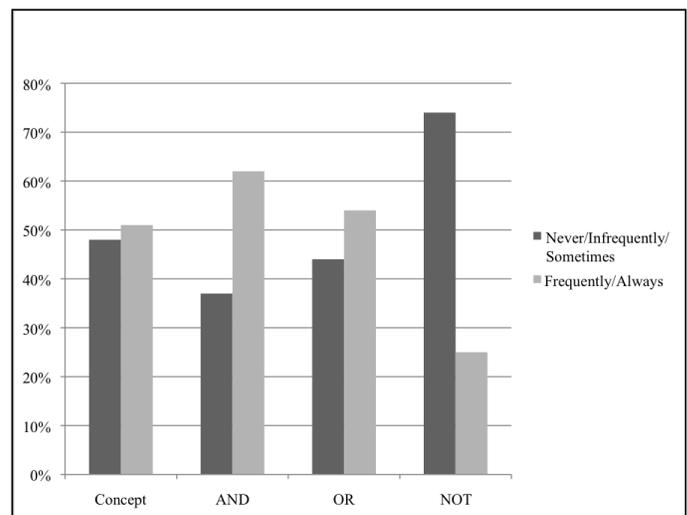
Two open-ended questions were included in the survey. The first question, "In what circumstances should librarians teach Boolean operators and why (e.g., always, just in credit-bearing courses, just to graduate students, in specific disciplines, with specific databases, etc.)?" elicited a variety of answers, from which several themes emerged. Nine respondents stated that it is most appropriate to teach Boolean operators to graduate or advanced students; however, one respondent gave the opposite answer, stating that she assumed graduate students already

know Boolean operators, and thus the information is repetitive. Seven respondents cited the need to teach Boolean operators with specific databases. Six respondents noted insufficient time as the reason they do not teach Boolean operators; many of these responses also indicated a need to cover more important material such as evaluation of sources. Three respondents explained that they only teach Boolean operators when teaching keyword searching. Three respondents attested to only teaching Boolean operators if and when a specific search needs connective logic. Although not directly asked, it is worthwhile to note that eight respondents specifically stated that they do not use the term 'Boolean' when teaching Boolean logic and/or Boolean operators.

The second open-ended question included in the survey asked respondents to "Please share your thoughts on database vendors automatically inserting Boolean operators in search strings, instead of requiring the end-user to insert appropriate Boolean operators." In answering this question 44% of respondents indicated a strong positive reaction to databases vendors automatically inserting Boolean operators into search strings, while 21% indicated a strong negative reaction to this phenomena, and 35% did not respond with a strong feeling either way. There was no clear correlation between length of career as a professional academic librarian and strong reactions to database vendors automatically inserting connective logic into end-user searches. Four respondents that have strong positive feelings for database-inserted search logic asserted that these positive feelings were only true when conjunction operators are introduced into the search strings; however, automatic insertions of disjunction operators are viewed negatively. One respondent put it this way:

> I like it when databases assume AND. I am not okay with databases that assume OR… I tell the student that… [this] type of search means, 'We feel really terrible that the computer did not find anything on your topic. To make US feel better, we have given you this large, irrelevant search instead.

CONCLUSIONS

This is a pilot research project, and results cannot be generalized to a greater population. The sample of respondents that teach library research credit-bearing courses is too small to draw any conclusions from, other than that it is clear not all professional academic librarians believe it is worthwhile to teach Boolean logic and connective operators even in settings where they have longer than the traditional one-shot to teach library research skills and concepts. It is also clear that within this population of professional academic librarians teaching library instruction one-shots there exist a range of beliefs concerning the need to teach Boolean logic and operators, as well as a wide variety of current practices. This research presents only a small snapshot of one side of a coin – the professional beliefs and practices of academic librarians; future research is necessary to determine what instruction students actually need in Boolean logic and Boolean operators to be successful in their research pursuits. In the past learning Boolean logic

and connective operators was essential to creating appropriate search strategies that retrieved relevant results. Today, database vendors automatically insert Boolean operators and continually refine search algorithms for relevancy rankings to retrieve better results for end users. The implications of these research findings are unclear. While undergraduates may continue to find the information sources needed for research and support the academic arguments they make, graduate students and faculty more often search for information on highly refined ideas and concepts. A new model may be needed for graduate student and faculty searching. This model may be based on greater instruction, but it may be that we, as a profession, move once again towards mediated searching for patrons who need more sophisticated search strategies.

REFERENCES

Association of College and Research Libraries. (2000). Information literacy competency standards for higher education. Chicago, IL: American Library Association. Retrieved from http://www.ala.org/acrl/sites/ala.org.acrl/files/content/standards/standards.pdf

Bolner, M. S., & Poirier, G. A. (2007). *The research process: Books & beyond* (4th ed.). Dubuque, IA: Kendall-Hunt.

Coffmann, S. (2012). The decline and fall of the library empire. *Information Today, 20*(3). Retrieved from http://www.infotoday.com/searcher/apr12/Coffman--The-Decline-and-Fall-of-the-Library-Empire.shtml

Ercegovac, A. (2008). *Information literacy: Search strategies, tools & resources for high school students and college freshmen* (2nd ed.). Columbus, OH: Linworth.

George, M. W. (2008). *The elements of library research: What every student needs to know*. Princeton, NJ: Princeton.

Hacker, D., & Fister, B. (2010). *Research and documentation in the Electronic Age* (5th ed.). Boston, MA: Bedford/St. Martin's.

Hock, R. (2010). *The extreme searcher's internet handbook* (3rd ed.). Medford, NJ: CyberAge Books.

Lyman, P., & Varian, H. R. (2000). Reprint: How much information? *Journal of Electronic Publishing, 6*(2). doi: 10.3998/3336451.0006.204.

Mann, T. (2005). The Oxford guide to library research (3rd ed.). Oxford, UK: Oxford.

O'Gormann, J. (2012). Reference products and services: Historical overview and paradigm shift. In S. Polanka (Ed.) *E-reference context and discoverability in libraries: Issues and concepts* (pp. 1-10). Hershey, PA: Information Science Reference.

Palmquist, M. (2006). *The Bedford researcher* (2nd ed.). Boston, MA: Bedford/St. Martin's.

Quaratiello, A. R., & Devine, J. (2011). *The college student's research companion: Finding, evaluating, and citing the resources you need to succeed* (5th ed.). New York, NY: Neal-Schuman.

Tensen, B. L. (2007). *Research strategies for a digital age* (2nd ed.). Boston, MA: Thomson Wadsworth.

Tyler, J. K., & Brahmi, F. A. (1993). The effect of a local area MEDLINE network on online end-user and mediated searching: An update. *Medical Reference Services Quarterly, 12*(4), 1-6.

Wiegand, W. A., & Davis, D. G. (1994). *Encyclopedia of library history.* New York, NY: Garland.

AUTHOR'S NOTE

Although this research project originated from a discussion of controlled vocabulary in bibliographic instruction, the research presented in this paper concerns Boolean operators in bibliographic instruction. The focus of the research changed as the concept of Boolean operators in bibliographic instruction is more basic than even controlled vocabulary; thus, it seemed logical to begin with research on Boolean operators.