

Handling Spelling Errors in Online Catalog Searches

Karen M. Drabenstott and Marjorie S. Weller

The purpose of this paper is to add to our understanding and knowledge of spelling errors in online catalog searches based on empirical studies of spelling errors in online catalog searches and suggest ways in which systems that detect such errors should handle the errors that they detect. One study of spelling errors in online catalog searches involved a categorization of user queries for subjects that were extracted from four university libraries' online catalog transaction logs. The results of the analysis demonstrated that less than 6% of user queries that match the catalog's controlled and free-text terms contain spelling errors. This percentage did not account for spelling errors in user queries that failed to match the catalog's controlled and free-text terms, because of the difficulty of verifying certain terms and phrases and of collection failure. The results of a related study involved user responses to an experimental online catalog that detected possibly misspelled words. While the majority of users corrected misspelled query words, a sizable proportion made an action that was even more detrimental than the original misspelling; for example, they added another word or phrase to the query in addition to the misspelled word. This paper concludes with three recommendations for improvements to online catalogs to assist users in the correction of misspelled query words and the detection of queries that fail due to collection failure.

Since the introduction of online catalogs in the early 1980s, librarians, system designers, and researchers have had a very accurate record of users' subject and known-item access points in the form of transaction logs. Dozens of researchers with varying intentions have studied the access points in these logs, especially access points that failed to produce retrievals. Some researchers merely described the subject and known-item access points that users entered into online catalogs, and others constructed rather elaborate schemes for categorizing

access points that were successful or unsuccessful at producing retrievals. One recurring problem that prevents the retrieval of bibliographic records is the occurrence of spelling errors in online catalog access points. Summing up our knowledge about spelling errors, we know that users make spelling errors; such errors are not very common in online catalog searches, but they do result in searches that fail to yield retrievals; and systems can be programmed to detect spelling errors in user-entered access points.

KAREN M. DRABENSTOTT is Associate Professor, School of Information, University of Michigan, Ann Arbor (e-mail: karen.drabenstott@umich.edu); MARJORIE S. WELLER is Programmer Analyst, Medical Center Information Technologies, University of Michigan, Ann Arbor (e-mail: mweller@m.imap.itd.umich.edu). Manuscript received December 15, 1995; accepted for publication February 9, 1996.

The purpose of this paper is to add to our understanding and knowledge of spelling errors in online catalog searches based on empirical studies of spelling errors in online catalog searches and suggest ways in which systems that detect such errors should handle the errors that they detect.

LITERATURE REVIEW

Many researchers have used transaction logs to study actual public use of computer-based retrieval systems. In his extensive review of transaction logs, Peters (1993) places studies of actual public use into nineteen different categories (e.g., errors, zero hits, missed opportunities, failures and their causes, extent of match studies, and user persistence).

Definitions of what constitutes a spelling error vary from transaction-log study to transaction-log study. Spelling errors could include nonlegitimate queries, such as random configurations, data-entry errors, and graffiti. In eight studies (Markey 1984, 66; Henty 1986, 48; Carlyle 1989, 51; Lester 1989, 172; Peters 1989, 270; Hunter 1991, 400; Zink 1991, 53; and Drabenstott and Vizine-Goetz 1994, 161), authors reported that a small proportion of the terms users entered into online catalogs were not legitimate subject queries. These included random configurations (///, HJNMV) and data-entry errors. Such activity could have been exploratory, accidental (e.g., leaning on the keyboard), or indicative of the frustration users were experiencing with their ongoing search. Such activity occurs with much less frequency than the entry of legitimate subject queries: for example, such activity was only 0.4% of the random sample of subject queries extracted from transaction logs at Northwestern University (Lester 1989, 172).

Some researchers consider the inclusion of punctuation in queries to be a spelling error (Henty 1986, 50; Walter 1987, 78; Lester 1989, 184; Drabenstott and Vizine-Goetz 1994, 173). Examples of punctuation occurring in user queries are possessive forms with an apostrophe, acronyms with periods between letters, hyphenated words and phrases, and inverted phrases entered with an intervening comma. Another type of spelling error is the entry of

phrases without intervening spaces (Henty 1986, 48; Jones 1986, 6; Walter 1987, 76). Words in the resulting phrase may or may not be spelled incorrectly.

The most common types of spelling error involve substitution, insertion, transposition, or omission of one or more letters in words. Substitution errors result in the substitution of one character for another, e.g., "lyprosy" instead of "leprosy." Queries bearing insertion errors contain extra letters, e.g., "peducation" instead of "education." Transposition errors result in two or more characters being reversed, e.g., "medeival" instead of "medieval." Omission errors occur when one or more characters are left out of the word, e.g., "lanuage" instead of "language." Nine transaction-log studies reported these kinds of common spelling errors (Markey 1984, 66; Henty 1986, 48; Jones 1986, 4; Walter 1987, 76; Lester 1989, 197; Peters 1989, 170; Hunter 1991, 400; Zink 1991, 53; and Drabenstott and Vizine-Goetz 1994, 175).

Researchers typically compare user queries with the words and phrases used in controlled vocabularies. Because certain characteristics of user queries prevent them from being exact matches of controlled vocabulary terms, researchers sometimes consider such queries as misspellings. For example, misspellings could include user queries that are singular forms of plural controlled vocabulary terms and vice versa, e.g., "mosquito" instead of the subject heading "Mosquitos" (Carlyle 1989, 44), or abbreviated forms of words or phrases, e.g., "20th century" instead of the subject heading "Twentieth century" (Henty 1986, 48; Walter 1987, 76; Carlyle 1989, 44).

Although user-assisted spelling-detection and -correction algorithms are commonplace in today's word-processing programs, such capabilities are not standard in online catalogs. An early catalog—BACS at Washington University—featured Soundex for spelling correction. Several versions of the experimental Okapi online catalog have featured user-assisted spelling detection and correction. Walker and Jones (1987, 76–77, 151) compared two versions of Okapi that featured

two slightly different user-assisted spelling-detection and -correction algorithms. In one version, possible misspellings were detected using a Soundex algorithm, users were informed of the possibly misspelled word, and one word was suggested as a replacement; users also were given the option to enter a new, different, or corrected query. The second version was the same as the first except that users were prompted to enter a new, different, or corrected word for the possibly misspelled word detected by Okapi. The researchers concluded that the former system handled 78% of cases well compared to the 64% of cases that the latter system handled well.

MISSPELLINGS IN USER QUERIES

Misspellings in user queries—the focus of this paper—came from data sets generated in two separate but related sponsored-research projects. The first project—titled “Enhancing a New Subject Access Design to Online Catalogs”—was supported by the OCLC Online Computer Library Center, Inc., Library and Information Science Research Grant Program (Drabenstott 1994). We obtained transaction logs from the online catalogs of Syracuse University, the University of California, Los Angeles (UCLA), the University of Kentucky, and the University of Michigan, extracted a total of about two thousand user queries for subjects from the logs, and performed a manual analysis of these queries. The manual analysis required us to categorize user queries according to the types of elements present in them (i.e., topical subjects, corporate names, geographic names, personal names, and combinations of two or more elements), develop subcategories of queries corresponding to the extent to which they matched subject headings and other subject-rich terms in bibliographic records, and identify queries that were neither matches of subject headings nor other subject-rich terms in bibliographic records. The results of the manual analysis demonstrated the extent to which users entered subject queries bearing misspellings into online catalogs.

The second project—titled “Testing a

New Subject Access Design to Online Catalogs”—was supported by the Department of Education’s College Library Technology and Cooperation Grants (Drabenstott and Weller 1995). The purpose of this research project was to test a new subject-access design. This design featured an online catalog that had a wide range of subject-searching capabilities and search trees to govern the system’s selection of searching capabilities in response to user queries. The system asked users to differentiate between subject queries bearing personal names and all other subject queries. On their own, the search trees then determined the extent to which user queries matched subject headings and other subject-rich terms in bibliographic records. This machine-based analysis resulted in the selection of a subject-searching approach that was likely to produce useful retrievals in response to user queries. Failure to effect a match between queries and the catalog’s vocabulary sometimes meant that the query word or words were misspelled. The experimental online catalog reported such queries to users and asked them to check their queries for possible spelling errors. The results of this interaction between system and users demonstrated how users would respond to an online catalog that assisted them in detecting misspelled queries.

The research questions addressed in this paper are: (1) How prevalent are misspellings in user queries for subjects? and (2) How do users respond to online catalogs that detect possible spelling errors in their queries for subjects?

For the analyses described in this paper we considered the following to be spelling errors: (1) substitution, (2) insertion, (3) transposition, (4) omission of one or more letters or spaces in words, and (5) run-on words missing one or more spaces.

PREVALENCE OF SPELLING ERRORS IN USER QUERIES

CATEGORIZING SUBJECT QUERIES EXTRACTED FROM TRANSACTION LOGS

A research team at the University of Michigan selected the initial queries users

entered in subject searches from the four libraries' transaction logs. We chose initial queries because subsequent queries might have been unnecessary if catalogs responded to initial queries with useful retrievals. Queries were selected from online catalog terminals searched exclusively by library patrons.

We categorized queries by the type(s) of elements present in them: (a) topical subjects, (b) corporate names, (c) geographic names, (d) personal names, and (e) combinations of two or more elements (a through d). We then subcategorized categorized queries using the same series of decisions that an online catalog that was programmed with search trees would make.

Search trees hold much promise for assuming the burden of determining which subject-searching approach is likely to produce useful information for user queries. The designers of the Okapi experimental online catalog first defined search trees as "a set of paths with branches or choices, which enables the system to carry out the most sensible search function at each stage of the search" (Mitev, Venner, and Walker 1985, 94). The search trees they implemented in Okapi evolved through a process of discussion and trial and error and placed more emphasis on searching the titles than the subject headings in Okapi's catalog records because only half of these records contained subject headings (Mitev, Venner, and Walker 1985).

Some operational online catalogs have subject-searching routines that resemble search trees. For example, the online catalog of the University of Illinois at Urbana-Champaign responds to user queries for subjects with keyword searches of assigned subject headings. When users terminate searches, the system prompts them to continue and gives the results of a title-keyword search (Hildreth 1989). The Illinois online catalog always performs keyword searches of subject-heading fields before title-keyword searches because the former consumes fewer system resources than the latter.

The search trees that we used to subcategorize categorized queries were the

result of the empirical study of the subject terms users entered into online catalogs (Drabenstott and Vizine-Goetz 1994). The empirical study demonstrated that the subject terms users entered into online systems possessed certain characteristics that revealed the subject-searching approaches most likely to succeed in producing assigned subject headings and bibliographic records on the topics users seek. Examples of such characteristics were the number of words in user queries, the extent to which user queries matched controlled vocabulary terms, and their ability to produce retrievals in response to certain subject-searching approaches.

SEARCH-TREE SUBCATEGORIES

Drabenstott and Vizine-Goetz (1994) discussed search trees in depth and provided flowcharts depicting search-tree decision points; thus, only a brief description of search-tree categories is given here. The first step was to segregate user queries containing personal names from user queries that did not contain personal names. The former queries were subjected to analyses that were different from the analyses performed on the latter queries. These latter queries were candidates for the exact-approach subcategory. To be placed in this category, user queries were compared with subject headings printed in the *Library of Congress Subject Headings (LCSH)* and with subject headings that could be formulated using subject headings printed in *LCSH* and subdivided by geographic, topical, and period subdivisions. On occasion, some manipulation would be necessary to effect an exact match. For example, matches were effected by the following: (a) ignoring capitalization, (b) removing punctuation, (c) removing stop words, (d) normalizing word order, (e) ignoring spelling, and (f) combinations of categories (a) through (e). In the event an exact match was made, no additional analysis of the query was done.

Queries were then given to a search tree that favored the alphabetical-approach subcategory. If queries matched a longer unsubdivided subject heading, they met the criteria for placement in the

alphabetical-approach subcategory. On occasion, some manipulation would be necessary to effect an alphabetical match. For example, matches were effected by the following: (a) ignoring capitalization, (b) removing punctuation, (c) ignoring spelling, and (d) combinations of categories (a) through (c). In the event an alphabetical match was made, no additional analysis of the query was done.

Remaining one-word queries were studied to determine whether title-keyword searches in operational online catalogs would produce retrievals. If title-keyword searches failed to produce retrievals, the query was probably misspelled or the result of collection failure.

Queries composed of two or more words that did not meet criteria for the exact approach or alphabetical approach remained. We performed two general types of keyword searches in operational online catalogs to produce retrievals: (1) keyword-in-heading searches through keyword-in-main-heading searches and keyword-in-subdivided-heading searches and (2) keyword searches through title-keyword searches, keyword-in-subject-heading-fields searches, and keyword-in-record searches. The order of keyword searches was important. Retrievals produced through the first few approaches should have been more precise than retrievals produced using the last two approaches because a single field (subject headings or titles) was searched.

Queries bearing personal-name elements were submitted to search trees requiring a different set of decisions. We began by differentiating personal-name queries bearing topical and other types of elements from queries bearing only name elements. We then tried to effect matches of the former with words in subdivided subject headings. That is, we tried to effect matches using the keyword-in-subdivided-heading search in operational online catalogs. If this failed, we performed keyword-in-record searches for all elements in user queries. If this also failed to produce retrievals, we ignored all elements except the personal-name elements in queries and used one or more remaining personal-name elements in

user queries to effect a match using the alphabetical approach.

Basic subcategories of user queries were the following:

- Exact matches (excluding queries with personal-name elements)
- Alphabetical matches (all queries)
- Keyword-in-heading matches (all queries)
- Keyword-in-record matches (all queries)
- Nonmatches (none of the above four subcategories, excluding queries with personal-name elements)

CATEGORIZED INITIAL ACCESS POINTS

A total of 1,919 initial access points in subject searches were extracted from the transaction logs of online catalogs at Syracuse University (571 access points), UCLA (511 access points), University of Kentucky (418 access points), and the University of Michigan (419 access points). The total percentages of types of initial access points across all four libraries are summarized in figure 1.

Overall, about 3 of every 5 queries contained only topical elements. Personal names accounted for 11% of user queries. The most frequent multielement query contained topical and geographic elements and represented about 8% of user queries for subjects. Nonlegitimate queries numbered 203; these were expletives, gibberish, explicit sex terms, known-item searches, and accounted for 10% of user queries. When nonlegitimate queries were discarded from subsequent analyses, a total of 1,716 subject queries were analyzed.

At all four data-collection sites, the majority of user queries for subjects were topical subjects. Users searching the online catalogs at Syracuse, Kentucky, and Michigan entered large percentages of subject queries for personal names. Subject queries for personal names that UCLA users entered were actually entered incorrectly. ULCA required users to use the system's Find Name or Browse Name commands to search personal-name queries rather than its Browse Subject command. Users searching online catalogs at UCLA and Michigan entered

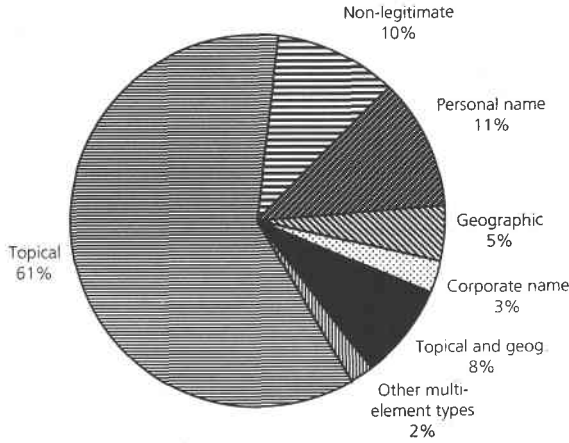


Figure 1. Types of Initial Access Points

large percentages of subject queries bearing both topical and geographic-name elements. Multiple-element queries—that is, queries bearing topical elements and one or more other element types—represented between 2% and 17% of user queries for subjects.

SUBCATEGORIZED INITIAL ACCESS POINTS

Exact Matches

Our analysis of user queries for subject generally began with a test to determine whether they were exact matches of con-

trolled vocabulary terms. Of the total of 1,716 legitimate subject queries, 832 queries (48.5%) met the criteria for exact matches (see table 1).

Of the 832 exact matches, most (653, or 78.5%) were queries for topical subjects. Spelling errors occurred in queries that were exact matches more frequently than in queries that were normalized matches. Spelling errors occurred in all types of categorized queries—queries bearing topical elements, geographic-name elements, corporate-name elements, and a combination of topical and geographic-name elements. Overall, spelling errors occurred in 5.8% of exact

TABLE 1
TYPES OF EXACT MATCHES

Type of Exact Match	Total (N=832)	Topical (N=653)	Geographic (N=79)	Corporate (N=40)	Topical- Geographic (N=60)
Exact	62.9	67.7	65.8	17.5	36.7
Exact, spelling error	4.3	4.1	8.9	5.0	0.0
Exact, reference	13.0	13.3	2.5	37.5	6.7
Exact, spelling error, reference	0.8	0.5	0.0	10.0	0.0
Normalized	16.0	12.1	20.3	25.0	46.6
Normalized, spelling error	0.7	0.5	0.0	2.5	3.3
Normalized, reference	2.3	1.8	2.5	2.5	6.7
Total	100.0	100.0	100.0	100.0	100.0

TABLE 2
EXAMPLES OF MISPELLED
EXACT MATCHES

User Query	Matching Subject Heading or Reference
catholic churchu	Catholic church
viet nam	Vietnam
psibocycin	Psilocybin
guadalupe	Guadaloupe
austrailia	Australia
austrialalia	Australia
syracuse univeristy	Syracuse University
phptpgraphy	Photography
3therapy	Therapy, <i>see</i> Therapeutics

matches. Examples of misspelled user queries that were exact matches of subject headings or references are listed in table 2.

Alphabetical Matches

Remaining queries were tested to determine whether they were alphabetical matches of controlled vocabulary terms. Of the total of 1,716 legitimate queries,

155 queries (9%) met the criteria for alphabetical matches (see table 3).

Except for queries bearing topical elements only, alphabetical matches were pretty rare. Spelling errors occurred in three of the four types of categorized queries—queries bearing topical elements, geographic-name elements, and a combination of topical and geographic-name elements. Overall, spelling errors occurred in 5.2% of alphabetical matches. Examples of misspelled user queries that were alphabetical matches of subject headings or references are listed in table 4.

Keyword-in-Heading Matches

Remaining queries were tested to determine whether they were keyword-in-heading matches of controlled vocabulary terms. Of the total of 1,716 legitimate queries, 98 queries (5.7%) met the criteria for keyword-in-heading matches (see table 5).

User queries that were keyword-in-heading matches were divided between queries bearing topical elements only and queries bearing a combination of topical and geographic elements. The majority of keyword-in-heading matches were matches of subdivided subject headings. Only about one-quarter of matches were

TABLE 3
TYPES OF ALPHABETICAL MATCHES

Type of Alphabetical Match	Total (N=155)	Topical (N=130)	Geographic (N=7)	Corporate (N=10)	Topical- Geographic (N=8)
Two or more words in heading	20.0	16.2	14.3	50.0	50.0
Two or more words in reference	4.5	4.6	0.0	10.0	0.0
Two or more words in reference, spelling error(s)	1.9	1.5	0.0	0.0	12.5
One word in heading	23.2	21.5	42.9	40.0	12.5
One word in reference	23.9	26.2	14.3	0.0	25.0
One word in heading, spelling error(s)	2.6	1.5	28.5	0.0	0.0
Less than one word in heading	18.7	22.3	0.0	0.0	0.0
Less than one word in reference	4.5	5.4	0.0	0.0	0.0
Less than one word in heading, spelling error(s)	0.7	0.8	0.0	0.0	0.0
Total	100.0	100.0	100.0	100.0	100.0

TABLE 4
EXAMPLES OF MISPELLED ALPHABETICAL MATCHES

Query	Matching Heading or Reference	Alphabetical-Match Type
carribean	Caribbean literature (French)	One word in heading, spelling error
chernoyble	Chernobyl Nuclear Accident, Chernobyl, Ukraine	One word in heading, spelling error
crcreative	Creative ability	One word in heading, spelling error
oorientalism	Orientalism in art	One word in reference, spelling error

TABLE 5
TYPES OF KEYWORD-IN-HEADING MATCHES

Type of Keyword-in-Heading Match	Total (N=98)	Topical (N=53)	Geographic (N=3)	Corporate (N=0)	Topical-Geographic (N=42)
Main heading	24.5	34.0	0.0	0	14.3
Subdivided heading	66.3	62.2	100.0	0	69.0
Subdivided heading, spelling error	9.2	3.8	0.0	0	16.7
Total	100.0	100.0	100.0	0	100.0

matches of main, unsubdivided subject headings. A little under 10% of keyword-in-heading matches involved spelling errors, and all were connected with matches of subdivided headings. Examples of misspelled user queries that matched subdivided subject headings are "south africa and te church" "2slave religion," and "pre-school testinhg."

Keyword Matches

Remaining queries were tested to determine whether they were keyword matches of controlled vocabulary terms. Of the total of 1,716 legitimate queries, 290 queries (16.9%) met the criteria for keyword matches (see table 6).

Over 85% of keyword matches were title matches. Such matches would provoke systems governed by search trees to respond to matches with title-keyword searches. Less than 4% of keyword matches would provoke systems to respond with keyword searches of subject heading fields. A little over 10% of keyword matches would result in keyword-in-

record searches. Spelling errors occurred in 5.5% of queries that were keyword matches. Examples of such misspelled queries were "development appropriate practiced," "socialization and sosiobiology," "critical succsess factors," and "tetropods."

Nonmatches

A set of 90 queries for topical subjects remained. These queries failed to meet the criteria for exact, alphabetical, keyword-in-heading, and keyword matches. Of these 90 queries for subjects generally, 71 queries bore topical elements only, 8 bore corporate-name elements only, and 11 bore a combination of topical and geographic elements. The reasons these queries were not matches of subject headings, words in titles, or words in other subject-rich fields of bibliographic records sometimes involved spelling. For example, the following nonmatching queries contained spelling errors:

- flamability standards
- black playwrightes
- federalism and jefersonianism

TABLE 6
TYPES OF KEYWORD MATCHES

Type of Keyword Match	Total (N=290)	Topical (N=249)	Geographic (N=3)	Corporate (N=1)	Topical- Geographic (N=37)
Two or more title words	73.5	73.9	100.0	100.0	67.6
Two or more title words, spelling error	4.1	4.8	0.0	0.0	0.0
One title word	7.2	8.0	0.0	0.0	2.7
One title word, spelling error	0.7	0.8	0.0	0.0	0.0
Words in subject heading fields	3.8	2.8	0.0	0.0	10.8
Words in subject-rich fields	10.0	8.9	0.0	0.0	18.9
Words in subject-rich fields, spelling error	0.7	0.8	0.0	0.0	0.0
Total	100.0	100.0	100.0	100.0	100.0

- rome oly
- transvesticism
- severely handicapped students

If some of these queries (e.g., "rome oly," "severely handicapped students," "transvesticism") were submitted to truncation, they might produce retrievals because truncation would be forgiving about spelling errors—that is, truncation would eliminate the misspelling from the word and a correctly spelled stem would remain. Other queries (e.g., "flamability standards," "black playwrights," "federalism and jefersonianism") might require both truncation and spelling correction to produce retrievals. The latter queries might not result in retrievals because they are too specific. Perhaps subject searches of journal article abstracts, back-of-the-book indexes, or tables of contents might produce retrievals, but these literary elements are seldom indexed in online catalog databases.

It was difficult for us to quantify spelling errors in nonmatching queries because we were unable to verify certain terms and phrases. Examples were "neander valley," "rayonnant architecture," "gnatting," "race g," "psychosis icu," and "cremastogaster pilosa." Much reference work and discussion with colleagues en-

abled us to verify the query "l'arche"—an international organization that assists mentally challenged adults—long after the project was completed.

Many nonmatching queries were spelled correctly. Examples are:

- smoking woman
- keystone corporation
- french occupation in chad
- luria-nebraska assessment battery
- toshiba affair

A combination of a number of techniques (e.g., truncation, matches on fewer than all words in queries) would probably result in matches that would lead to retrievals. As a last resort, searches of journal article abstracts, back-of-the-book indexes, or tables of contents might produce retrievals because the subjects represented by these queries are too specific to be treated in full-length books, monographs, and journal titles.

SPELLING ERRORS AND MATCHES

Spelling errors in 5.9% of 1,375 queries for subjects generally prevented exact, alphabetical, keyword-in-heading, and keyword matches. Spelling errors were not pervasive in a particular match type. Generally, spelling errors occurred in between

5% and 6% of exact, alphabetical, keyword-in-heading, and keyword matches.

Although spelling errors in an additional 90 queries for subjects generally occurred in only a fraction of these queries, systems would have a difficult time finding them on their own because retrievals would be possible only after spelling was corrected and the system performed a second matching technique (e.g., truncation, searching for fewer than all words in queries). Spelling errors would also be difficult to detect because the system's failure to produce retrievals might be due to collection failure. Searches that are more comprehensive or larger than library cataloging databases might provide more detail and greater depth about an item's subject matter, e.g., tables of contents, back-of-the-book indexes, or full texts.

SUBJECT QUERIES FOR PERSONAL NAMES

Search trees for subject queries for personal names consider user queries bearing personal-name and topical elements as candidates for keyword-in-heading matches or keyword matches. Failure to produce matches for these two keyword matches would result in the omission of topical elements and the submission of name elements only to alphabetical matching. Of the total of 251 queries that contained personal-name elements, only 32 queries contained a combination of personal-name and topical elements; thus, these queries were submitted to keyword-in-heading matches or keyword matches. Three queries contained spelling errors that prevented keyword matches, viz. "skinner and siblings," "delacroix and colr," and "clarence darrow's relegious views." Omitting the misspelled topical elements from each query left one or more personal-name elements that could be used by the alphabetical approach to find the appropriate location in an alphabetical list of personal-name subject headings where the personal-name elements of the query might be listed.

Search trees would treat the remaining 219 names bearing personal-name ele-

ments only by submitting them to alphabetical matching—that is, using the name elements to find the appropriate location in an alphabetical list of personal-name subject headings where the personal-name elements of the query might be listed.

Quantifying spelling errors in subject queries for personal names was irrelevant for two reasons. First, computer-based retrieval systems can perform matching techniques that forgive spelling errors. Second, some queries bore names that were impossible for us to verify, so we did not know whether such queries contained misspellings or named individuals for whom no monographic literature was available. When queries for personal names were misspelled, users might have found the desired personal-name heading rather quickly because the spelling error was toward the end of the name or the name stem was rather unique and there were likely to be few names beginning with the stem. Examples are the following misspelled queries for personal names:

- bosc hieronymus
- shakespear
- philoctetess
- aphropdite
- nanbnancy holt

Users who entered queries for the following misspelled names into systems that responded with an alphabetical list of personal-name subject headings would probably have to browse many, many lists to find the desired name quickly because the spelling error was at or toward the beginning of the name or the basic word stem was not unique and there were likely to be many names beginning with the stem:

- hitckcock
- farakan louis
- daili
- magrite rene
- lfeinbloom deb
- hypocrates
- nmarrow philip

We were unable to verify a total of fourteen names. Examples are:

- prosser walter lee
- steinway henry
- n schribner richard
- abrahams r d

- klagsburn

Search trees for personal-name queries are forgiving about spelling errors because they always respond with an alphabetical list of personal-name subject headings in the alphabetical neighborhood of user-entered, personal-name elements of queries. User perseverance and the uniqueness of personal-name elements in queries vis-à-vis personal-name subject headings in the alphabetical neighborhood of the desired name have a large stake in the outcome—that is, whether users find the desired names.

USING AN EXPERIMENTAL ONLINE CATALOG TO DETECT POSSIBLE MISSPELLINGS

For our second research question we report on the results of user responses to an experimental online catalog that detected possible misspellings in user queries for subjects. The experimental online catalog was developed in a research project titled "Testing a New Subject Access Design to Online Catalogs." This design featured an online catalog that had a wide range of subject-searching capabilities and search trees to govern the system's selection of searching capabilities in response to user queries. The search trees utilized matching techniques to determine the extent to which user queries matched subject headings and other subject-rich terms in bibliographic records. These techniques were the same as the techniques used in the matching study that is the subject of the first half of this paper. That is, the system asked users to differentiate their queries for subjects generally from their queries for personal names. Search trees then sought exact, alphabetical, keyword-in-heading, and keyword matches of subject headings or subject-rich fields of bibliographic records and responded with subject-searching approaches corresponding to the types of matches made. Search trees chose subject-searching approaches that were likely to produce useful retrievals in response to user queries; thus, they favored controlled vocabulary over free-text searching approaches. Failure to effect a match between queries and the

catalog's vocabulary sometimes meant that the query word or words were misspelled. The experimental online catalog reported such queries to users and asked them to check their queries for possible spelling errors. The results of this interaction between system and users demonstrated how users would respond to an online catalog that assisted them in detecting misspelled queries and, thus, answered our second research question, viz. "How do users respond to online catalogs that detect possible spelling errors in their queries for subjects?"

EXPERIMENTAL ONLINE CATALOG DEVELOPMENT

The experimental online catalog named ASTUTE (A Search Tree Underlying The Experiment) was developed by a project team at the University of Michigan to test the new subject-access design. The team programmed ASTUTE on a stand-alone Gateway 2000 486, 33 MHz, IBM-compatible microcomputer, with 8 megabytes of RAM and a VGA color monitor. The operating system was MS-DOS version 5.0. A dot-matrix printer and a mouse were attached to the microcomputer for use by ASTUTE project staff during development work and end users during on-line retrieval tests.

The databases of the ASTUTE experimental online catalog were created from two data sources: (1) Machine-Readable Cataloging (MARC) records for bibliographic data from the two participating libraries in selected subject areas of the Library of Congress Classification (LCC) and (2) MARC records for subject-authority data from the compact disc-based product CD/MARC Subjects distributed by the Library of Congress. The number and subject areas of MARC bibliographic records were:

1. Mardigian Library of the University of Michigan—Dearborn: 14,686 bibliographic records in Computer Science (QA76) and Technology (T-TX)
 2. Lilly Library of Earlham College: 11,976 bibliographic records in American History (E1-F1199)
- The ASTUTE project team did not

combine bibliographic records into a single database. Rather, the team used the two libraries' bibliographic records to create separate, searchable databases, one on computer science and technology for the University of Michigan-Dearborn (UM-D), and one on American history for Earlham College.

SUBJECT SEARCHING IN THE EXPERIMENTAL ONLINE CATALOG

We tested the retrieval effectiveness of the experimental online catalog with search trees by comparing its performance with the performance of an experimental online catalog in which subject-searching approaches were assigned at random. To accomplish this, we designed the ASTUTE experimental online catalog to feature two online catalogs: (1) the Blue System, in which search trees governed the system's selection of a subject-searching capability, and (2) the Pinstripe System, in which the system selected a subject-searching capability randomly. These systems were purposely designed to be very much alike to focus the attention of library patrons and staff on the retrieval of useful information in response to their queries. The Blue and Pinstripe Systems had virtually the same interfaces, and they accessed the same bibliographic and authority databases. Except for the Blue System's enhancement with the search trees, the two systems and their capabilities were the same.

Search trees exemplified the searching strategies used by expert search intermediaries. Intermediaries use controlled vocabulary because it yields relevant output. When controlled vocabulary is not available to express user queries, intermediaries conduct free-text searches of titles and abstracts to retrieve a few relevant records, review results to find relevant controlled vocabulary, and then incorporate such vocabulary into the ongoing search. The search trees performed in a similar manner. They invoked searching approaches that looked for matches of user queries in subject-heading fields of cataloging records before enlisting keyword-search approaches that looked for

matches in title fields or in a combination of title and subject-heading fields. Like the matching studies in the first half of this paper, search trees for queries for subjects generally effected exact matches, alphabetical matches, and keyword-in-heading matches, that is, matches of controlled vocabulary terms, before effecting keyword matches, that is, matches of free-text words and phrases in bibliographic records. Also, search trees for subject queries for personal names effected keyword-in-heading matches of name and topical elements in user queries before ignoring topical elements and displaying an alphabetical browsing list of personal-name subject headings in the alphabetical neighborhood of personal-name elements in user queries. Thus, decisions that the search trees made about responding to user queries with matches of subject headings and words in bibliographic records were very similar to the decisions that judges made about matching user queries in the matching study described in the first half of this paper.

DETECTION OF POSSIBLY MISPELLED WORDS IN USER QUERIES

The ASTUTE experimental online catalog did not feature automatic spelling correction. It did, however, inform users of query elements that failed to produce retrievals and suggest users check such elements for spelling errors.

The Blue System checked to determine whether each word in user queries for subjects generally was posted in its database following the system's failure to make matches through the exact and alphabetical approaches. It checked words in queries from left to right. An example is the user query "noegro peospirity in late 1920's." The system failed to find the word "noegro," informed the user of its failure, and suggested the user check spelling (see figure 2). The user corrected this word and submitted a new query to the system—"negro peospirity in late 1920's." The system failed to find the word "peospirity," informed the user of its failure, and suggested the user check spelling (see figure 3). After responding to several sys-

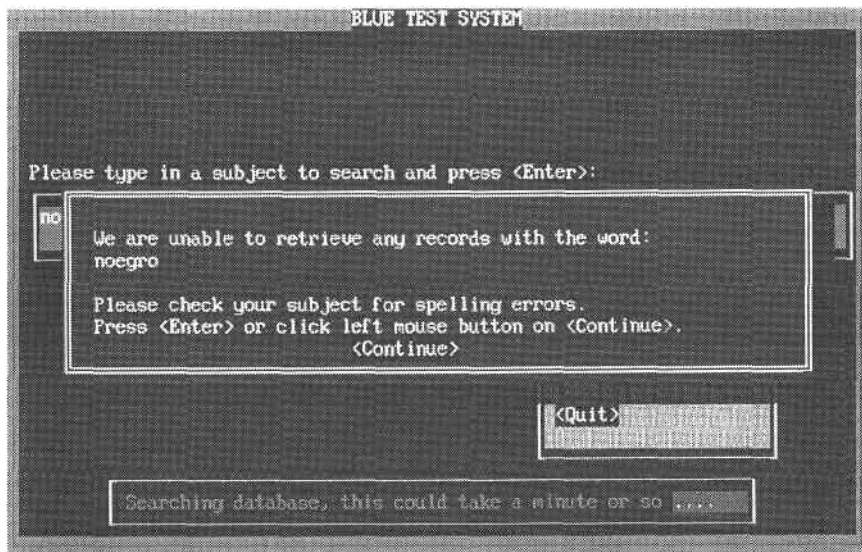


Figure 2. Informing Users of Possible Misspellings

tem prompts to check spelling, the user eventually entered a query that contained no spelling errors, viz., "negro prosperity in late 1920's."

When the Pinstripe System's random-selection algorithm called for the alphabetical approach, the system made no attempt to find possible spelling errors. When it called for keyword-in-subdivided-heading or keyword searches, the system performed the same error-checking routine as the Blue System. That is, it checked a keyword index to determine whether the individual words in queries were used in the database. It checked the words in queries from left to right, informed users of query elements that failed to produce retrievals, and suggested users check such elements for spelling errors.

Queries that the Blue and Pinstripe Systems identified as having possible spelling errors would be considered non-matches in the matching study that is the focus of the first half of this paper. This did not always mean that query words were misspelled. Automatic truncation or

matches on fewer than every word in queries or a combination of these two techniques could have resulted in matches and, thus, retrievals. Also, collection failure could be the reason why ASTUTE failed to produce retrievals for words in user queries.

ADMINISTERING COMPARISON SEARCH EXPERIMENTS IN LIBRARIES

The ASTUTE project team transported the Gateway microcomputer bearing ASTUTE to the two data-collection sites—Mardigian Library at the University of Michigan-Dearborn and Lilly Library at Earlham College. The microcomputer was dedicated to use of the ASTUTE experimental online catalog. At UM-D, ASTUTE was located in a quiet study area of the library that was also near the computer science, engineering, and technology stacks. Thus, ASTUTE searchers would not have to go very far to access the library material they retrieved in their searches of the experimental online catalog. At Earlham College,

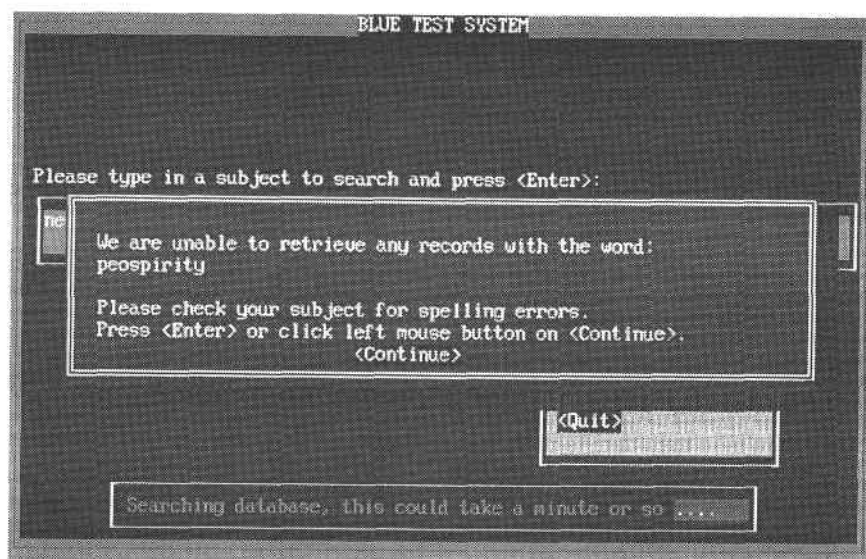


Figure 3. Checking Queries from Left to Right for Possible Misspellings

ASTUTE was located in the reference area of the library near the library's MARCIVE CD-ROM-based online catalog and other CD-ROM reference sources. Lilly Library reference staff were also nearby and directed patrons to ASTUTE when they felt patrons would find useful material in the system. At both libraries, signs were placed near ASTUTE to attract library patrons to use the system.

The ASTUTE experimental online catalog performed recruiting functions on its own. Introductory screens invited users to participate in the experiment; told users how to operate the keyboard and mouse, make selections, and print screens; and asked them to conduct a computer-based search on a topic of their own choosing in the system. ASTUTE told users it was logging their searches, relevance assessments to displayed titles, and responses to questions. Library users were entirely on their own to read screens, conduct searches, and answer questions.

The data-collection period at UM-D lasted five weeks, from March 12 to April 19, 1993. ASTUTE administered a total of

826 Comparison Search Experiments. At Earlham College, data collection lasted thirteen weeks, from February 23 to May 28, 1993. ASTUTE administered a total of 238 Comparison Search Experiments. Thirty-three of the total 1,064 search administrations involved library staff at the two participating libraries.

Interviewers were not present to monitor system use; consequently, we expected searches for topics that were not represented in the experimental online catalog. We also expected searchers to leave the experiment without completing the full search administration. To determine usable search administrations for submission to data analyses, the ASTUTE project team had to manually review searches and queries. Of the 1,064 search administrations, about half (528 of 1,064 administrations) were usable. About a third (34%) were unusable queries that were entered into the experimental online catalog's subject-searching capabilities for subjects generally. About three-quarters of these unusable subject queries were out-of-scope, that is, the bibliographic-record databases did not contain titles for

the requested topics. Most other unusable subject queries were characterized as playing or meaningless input, e.g., sex terms, expletives, blanks, one or more of the same letters, gibberish. About one-eighth (12%) were unusable queries for personal names. Some unusable queries for personal names were names that were out-of-scope, others were elements of known-item searches, and others were playing or meaningless input. Less than 5% of unusable searches were search administrations in which users completed one or more presearch questions, but they did not continue with their searches. These users probably walked away from the system, and it eventually reset itself to the introductory screen savers.

A large percentage (43%) of usable administrations of the Comparison Search Experiment were full administrations. Of the four partial-administration categories, the largest percentage (29%) contained the three complete events; unfortunately, users walked away before completing the postsearch questionnaire.

Details about individual search administrations are given in the final report of the project (Drabenstott and Weller 1995). Our focus in this paper is on users'

responses to the experimental online catalog's suggestions that their queries might be misspelled.

USER RESPONSES TO ASTUTE'S SUGGESTION OF MISPELLED WORDS IN USER QUERIES

The experimental online catalogs responded to 134 queries with the messages in figures 2 and 3 that informed users that their queries contained possible spelling errors. In table 7 we describe what users did next.

Large numbers of users entered queries on different topics. Examples of queries bearing unposted words and the queries users entered following the system's message informing them of a possible spelling error are listed in table 8. Words in italics were the unposted words that the experimental online catalog displayed to users for their correction.

Following the system's message informing them of a possible spelling error, large numbers of users entered the same query one or more times. Examples of such queries are "internet," "hovercraft," "androids," "barcode," "reinforce concrete," and "nanotechnology." Perhaps

TABLE 7
USER ACTIONS FOLLOWING SYSTEM MESSAGE REGARDING
POSSIBLE SPELLING ERRORS

User Actions	UM-D		Earlham	
	No.	%	No.	%
Entered query on different topic	34	27.2	1	11.1
Entered same query	27	21.6	0	0.0
Quit search	15	12.0	2	22.2
Corrected spelling	11	8.8	5	55.6
Entered same query minus unposted word	12	9.6	0	0.0
Entered same query and added new word(s)	11	8.8	0	0.0
Entered new query with same stem as previous query	7	5.6	0	0.0
Entered singular or plural form of previous query	7	5.6	0	0.0
Entered acronym or spelled it out	1	0.8	1	11.1
Total	125	100.0	9	100.0

such users reentered the same queries because they wanted to make absolutely sure that the system had no titles on these topics. Reentering such queries, users might have been saying to themselves, "There's got to be information on this topic in here somewhere."

Following the system's message informing them of a possible spelling error, fifteen users at UM-D and two users at Earlham quit searching.

A total of sixteen users corrected the misspellings in their queries (see table 9—words in italics were unposted words that the experimental online catalogs suggested to users were misspelled).

Of the total 134 queries in which the experimental online catalog detected unposted words, 28 queries actually contained misspelled words. Users corrected 16 of these queries. Examples of queries

that users did not correct were "electronics," "circuits," "assembly language," "cobal languages," and "e\\EIFFEL." In response to the system's message informing the latter users of a possible misspelling, some users quit searching, other users entered different queries, and still other users added new words or deleted the possibly misspelled words from queries.

A total of 23 users responded to the system message about possible misspellings by entering the same query minus the unposted word or by adding a new word to the same query (see table 10—words in italics were unposted words that the experimental online catalogs suggested to users were misspelled).

A handful of queries was placed in remaining categories. Examples of succeeding queries that had the same stem as preceding queries were "automanual" and "auto," "encoding" and "encode," "architectual design" and "architecture." Examples of succeeding queries that were singular or plural forms of preceding queries were "florida keys" and "florida key," "air bag" and "air bags," and "rotation of axis formula" and "rotation of axes formula." Two queries contained acronyms, "IWW" and "International Workers of the World" and "cobal languages" and "common business oriented language."

Frequently, the experimental system informed end users that their entered terms might be misspelled when, in fact, their entered terms were not posted in the database.

TABLE 8
SUCCEEDING QUERIES
ON DIFFERENT TOPICS

Possibly Misspelled Queries	Next Queries on Different Topic
<i>internet</i>	usenet
microcad <i>graphing</i>	database
<i>radiator</i> design	heat transfer
<i>zirconia</i>	chemistry
general <i>relativity</i>	computers
The z8 microcomputer	zilog microcomputers

TABLE 9
MISSPELLED QUERIES

Possibly Misspelled Queries	Corrected Queries
human power <i>vechicles</i>	human power vehicles
<i>communcations</i>	communications
<i>carsuspension</i> and handling	car suspension and handling
<i>chemistryy</i>	chemistry
<i>abolishionism</i>	abolitionism
c programming <i>lanuage</i>	c programming language
<i>monitring</i> performance of telephone operators	monitoring performance of telephone operators

TABLE 10
 QUERIES WITH WORDS
 ADDED OR DELETED

Original Queries	Added or Deleted Words in Subsequent Queries
<i>rotation</i> of axis	rotation of axis formula
probability and <i>statistics</i>	probability
<i>barcode</i>	2d barcode
<i>equilizer</i> filters	filters
<i>graphing</i>	microcad graphing
<i>cache</i> memory	memory
<i>internet</i>	internet network
<i>three-dimentional</i> dynamics	dynamics

INCORPORATING SPELLING ASSISTANCE IN ONLINE CATALOGS

Users enter subject queries that contain spelling errors. This is not an especially serious problem with respect to legitimate user queries for subjects generally because spelling errors occur in a little less than 6% of such queries.

Spelling is also not a serious problem with respect to legitimate user queries for personal names, especially in catalogs that respond to such queries with an alphabetical listing because users can browse alphabetical lists to find the desired names. It is, however, quite difficult for systems to distinguish on their own personal-name elements of user queries from topical and other types of elements. If users distinguish such elements for systems, systems can then use this knowledge to check the spelling of topical and other non-name elements and, as a last resort, respond to users with the results of an alphabetical search for the personal-name elements only of personal-name subject queries when users fail to correct misspelled non-name elements. It is difficult for systems to detect misspelled personal-name elements because of the many variants for even seemingly simple names, e.g., Smith, Smithe, Smidth, Smitt, Smitz, Smyth, or Smythe. The alphabetical approach that was the default response in ASTUTE's

Blue System to personal-name queries bearing personal-name elements only will help users whose personal-name queries are in the same alphabetical neighborhood as listed personal-name subject headings or especially persevering users who are willing to browse backward and forward for the desired name.

Despite the infrequency of spelling errors, such errors can completely derail the most routine subject search. Examples come from a search in which a user began by using the misspelled term "lyprosy" followed by 45 other access points that either retrieved material that was too broad (using queries such as "microbiology," "skin diseases," and "skin growth") or that failed to retrieve any material due to other spelling errors or collection failure, e.g., "lepors," "lyprosy" (entered multiple times), and "hansen's disease," and a search in which a user entered the misspelled query "mideival art" three times, received no guidance from the system as to the correct spelling of the misspelled query word "mideival," and then walked away. In view of these two users' behavior, we can speculate that neither user knew that the root of the problem was a misspelled query.

On one hand, we can continue to allow our online catalogs to fail our users in view of the infrequency of spelling errors. On the other hand, we can also make rather simple enhancements to our existing online catalogs to help users overcome misspelled queries. Here are three suggestions.

First, online catalogs should be equipped with search trees to place the burden of selecting a subject-searching approach in response to user queries on the system instead of on users. An empirical study of search-tree effectiveness demonstrated that the search trees were more effective in selecting a subject-searching approach that would produce useful information for the subjects users seek than users would select on their own (Drabenstott and Weller 1995). Search trees considerably reduce search-approach failures. These failures are the direct result of the failure of a particular search approach to retrieve useful retriev-

als in response to user queries. Search trees enlist all search approaches in a deliberate sequence that begins with controlled vocabulary approaches that are more likely than free-text approaches to retrieve relevant material. Search trees also include tactics that are intended to overcome spelling errors. For example, search trees check the individual words in non-name elements to determine whether they produce retrievals. While this tactic is intended to conserve system resources connected with keyword searching, it produces intermediary results that are useful to online catalog users, because if one or more words in a query that is submitted to keyword searching fail to produce retrievals, keyword and implicit Boolean searching will also fail. Thus, it makes sense for the system to report intermediary results to users so that they can decide what to do with the offending words. Another example is the use of alphabetical searching for subject queries bearing personal names only or a combination of personal-name and non-name elements that fails to produce retrievals for both name and non-name elements. The alphabetical approach gives users the opportunity to browse backward and forward in alphabetical lists to find the desired names.

Second, when systems are unable to produce retrievals for elements of user queries, they should inform users and suggest one or more correct spellings of the possibly misspelled word. In view of the popularity of word-processing programs that have such spelling-correction routines at the present time, online catalog users might come to expect such assistance from online catalogs.

Third, while computer-assisted spelling routines in online catalogs can help users and systems identify misspelled words, they cannot distinguish between words that fail to produce retrievals because of misspellings or collection failure. Online catalog indexes could be enhanced with words and phrases from dictionaries, subject-heading lists, thesauri, and various other specialized and authoritative subject vocabularies. When query words match unposted words in these vocabular-

ies, this would be an indication that the failure to produce retrievals was due to collection failure and not spelling. Systems could even use the knowledge of the match to suggest that users search a specialized database. For example, suppose that words in a user query matched words from a specialized dictionary or thesaurus in zoology that were not posted in the online catalog's database. The system could use this knowledge to suggest that the user search a specialized zoology database or a general science database that provides access to abstracting and indexing records to journal articles.

SUMMARY

The purpose of this paper is to add to our understanding and knowledge of spelling errors in online catalog searches based on empirical studies of spelling errors in online catalog searches and suggest ways in which systems that detect such errors should handle the errors that they detect.

An empirical study of spelling errors in online catalog searches involved a categorization of user queries for subjects that were extracted from four university libraries' online catalog transaction logs. The results of the analysis demonstrated that less than 6% of user queries that match the catalog's controlled and free-text terms contain spelling errors. This percentage does not account for spelling errors in user queries that fail to match the catalog's controlled and free-text terms. It was difficult for the researchers to quantify spelling errors in nonmatching queries because we were unable to verify certain terms and phrases. We concluded that a combination of a number of techniques (e.g., truncation, matches on fewer than all words in queries) would probably result in matches that would lead to retrievals. As a last resort, searches of journal article abstracts, back-of-the-book indexes, or tables of contents might produce retrievals, but few online catalogs index terms from these sources in their databases.

An empirical study of online catalog use tested a new subject-access design. This design featured an online catalog that

had search trees to govern the system's selection of searching routines in response to user queries. Search trees determined the extent to which user queries matched subject headings and other subject-rich terms in bibliographic records. This machine-based analysis resulted in the selection of a subject-searching approach that was likely to produce useful retrievals in response to user queries. Failure to effect a match between queries and the catalog's vocabulary sometimes meant that query words were misspelled. The experimental online catalog reported such queries to users and asked them to check their queries for possible spelling errors. The results of this interaction between system and users demonstrated that users responded in several different ways to an online catalog that assisted them in detecting misspelled queries. Some ways resulted in a successful search. For example, the system identified a misspelled word, the user corrected the spelling, and the system produced useful retrievals for the corrected query. Some ways resulted in an unsuccessful search. For example, the system identified a misspelled word, the user did not correct the spelling, and, instead, added another word or phrase to the query in addition to the misspelled word. The experimental online catalog detected a total of 134 queries in which words were possibly misspelled. Of these queries, only 28 queries contained misspelled words, and users corrected 16 of these queries. Many of the remaining 106 queries were not spelled incorrectly. Instead, they contained words that were not in the catalog's database and, thus, were queries that failed due to collection failure. Searches of more comprehensive databases or records that had more depth than records in library cataloging databases might produce useful retrievals for these queries.

We concluded with three recommendations to improve the responsiveness of online catalogs to user queries that may be marred by spelling errors. First, we recommended that online catalogs be equipped with search trees to place the burden of selecting a subject-searching approach in response to user queries on

the system instead of on users and, thus, reduce search-approach failures in subject searching. Search trees also utilize tactics that are intended to overcome spelling errors such as the alphabetical approach, which gives users the opportunity to browse backward and forward in alphabetical lists to find the desired names. Second, we recommended that systems be equipped with automatic spelling-detection routines that, at the very least, inform users of a possibly misspelled word or words. Third, we recommended that online catalogs be enhanced with tools and techniques to distinguish between queries that fail due to misspellings and collection failure.

In closing, we caution that spelling is not a serious problem in subject retrieval, but, unfortunately, a problem as simple as spelling can completely derail the most routine subject search. Users expect and have become accustomed to spelling-correction routines in off-the-shelf word-processing software. Isn't it time to provide them with spelling correction in online catalog searching?

WORKS CITED

- Carlyle, Allyson. 1989. Matching *LCSH* and user vocabulary in the library catalog. *Cataloging & classification quarterly* 10, no. 1/2: 37-63.
- Drabenstott, Karen M. 1994. *Enhancing a new design for subject access to online catalogs*. Ann Arbor, Mich.: School of Information and Library Studies, Univ. of Michigan.
- Drabenstott, Karen M., and Diane Vizine-Goetz. 1994. *Using subject headings for online retrieval: Theory, practice, and potential*. San Diego: Academic Pr.
- Drabenstott, Karen M., and Marjorie S. Weller. 1995. *Testing a new design for subject access to online catalogs*. Ann Arbor, Mich.: School of Information and Library Studies, Univ. of Michigan.
- Henty, Margaret. 1986. The users at the online catalogue: A record of unsuccessful keyword searches. *LASIE* 17, no. 2: 4-52.
- Hildreth, Charles R. 1989. *Intelligent interfaces and retrieval methods for subject searching in bibliographic retrieval systems*. Advances in library information technology 2. Washington, D.C.: Library of Congress.

Hunter, Rhonda N. 1991. Successes and failures of patrons searching the online catalog at a large academic library: A transaction log analysis. *RQ* 30: 395-402.

Jones, Richard. 1986. Improving Okapi: Transaction log analysis of failed searches in an online catalogue. *Vine* no. 62: 3-13.

Lester, Marilyn Ann. 1989. Coincidence of user vocabulary and *Library of Congress Subject Headings*: Experiments to improve subject access in academic library online catalogs. Ph.D. diss., University of Illinois at Urbana-Champaign.

Markey, Karen. 1984. *Subject searching in library catalogs: before and after the introduction of online catalogs*. Dublin, Ohio: OCLC.

Mitev, Nathalie, Gillian Venner, and Stephen Walker. 1985. *Designing an online public access catalog*. Library and information research report 39. London: British Library.

Peters, Thomas A. 1993. The history and development of transaction log analysis. *Library hi-tech* 11, no. 2: 41-66.

———. 1989. When smart people fail: An analysis of the transaction log of an online public access catalog. *Journal of academic librarianship* 15: 267-73.

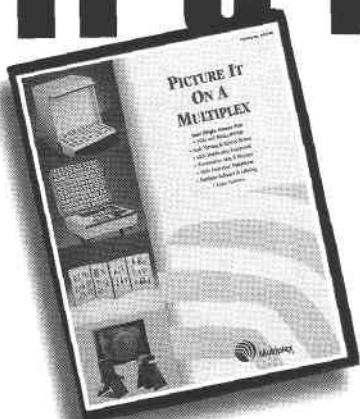
Walker, Stephen, and Richard M. Jones. 1987. *Improving subject retrieval in online catalogues; 1. Stemming, automatic spelling correction and cross-reference tables*. British Library Research Paper no. 24. London: British Library.

Walter, Dennis R. 1987. The user at the online catalogue: A record of unsuccessful keyword searches—another case study. *LASIE* 18, no. 3: 74-81.

Zink, Steven D. 1991. Monitoring user search success through transaction log analysis: The WolfPAC example. *Reference services review* 19, no. 1: 49-56.

IT'S HERE!

Your FREE Multiplex Catalog



Everything You Need For Convenient Slide Management:

- Slide and Media Storage
- Slide Viewing and Editing
- Archival Storage Products
- Presentation Aids & Displays
- Laser Pointers
- Assorted Accessory Items

Reserve your FREE copy today
FAX: 314-326-1716 or Call: 800-325-3350



Multiplex Display Fixture Company • 1555 Larkin Williams Road • Fenton, MO 63026-3008