

# The Effect of Misspellings on Information Retrieval in Online Public Access Catalogues

**Rebekah Willson**  
[rwillson@ualberta.ca](mailto:rwillson@ualberta.ca)  
University of Alberta

**Lisa M. Given**  
[lisa.given@ualberta.ca](mailto:lisa.given@ualberta.ca)  
University of Alberta

---

## Abstract

Spelling is an important literacy skill, crucial to successful searching in Online Public Access Catalogues (OPACs). Students at the University of Alberta, were given 4 search tasks. They were given a topic and asked to find as many items on that topic in the OPAC as possible. They were allowed to use any Internet resources of their choice to aid their search. Half of the participants were given easy-to-spell search terms (Easy Group), the other half difficult-to-spell (Difficult Group).

There were several factors that affected students' searching and information retrieval. The years of education increased the likelihood that participants would check spelling and gather background information on the topic. The frequency of Internet use also increased the likelihood that students would check spelling and use broader search fields. The difficulty of spelling level of the search topics affected: how participants searched for the term; whether or not they checked the spelling; whether or not they gathered background information; and the number misspellings and unsuccessful searches. The search behaviour of participants indicates that students search OPACs like they do the Internet. Despite some ineffective OPAC searching, most participants were successful in obtaining the correct spelling of the search terms, which was important because misspellings made in the OPAC led to no search results (with the exception of misspellings that were contained in the catalogue records). The results obtained have implications for including spell checkers in OPACs, cataloguing and information literacy instruction.

## 1. Introduction

While research has been done on how Online Public Access Catalogues (OPACs) are used, little research has been done on how spelling affects information retrieval. Spelling is a topic that has received little attention in library and information studies (LIS), though it affects both information retrieval and search behaviour. Correct spelling is important when searching library catalogues or online; however, many people see the use of spell checkers as rendering proper spelling irrelevant.

In addition to spelling there are many other factors that can affect searching, including understanding of how OPACs operate (Novotny 2004), unfamiliarity with library terminology and academic level of the searcher (Malliaris and Kyriaki-Manessi 2007). Information searching is a task that many find difficult. Students perform information searches on a regular basis. In addition to being required to perform searches, students are often asked to perform searches on topics about which they know little. Searching for unfamiliar topics can create problems when searching (e.g. Dewdney and Michell 1996).

This study will aid in understanding how university students search for information with difficult-to-spell search terms, about which they may know little, and what actions students take when misspellings render their searches unsuccessful. This information will contribute to a better understanding of how misspellings affect information retrieval during OPAC searches. It is hoped that with this understanding, users can be taught new techniques to better use OPACs. This research can also add to a growing body of knowledge about how OPACs can be designed to accommodate users.

### 1.1 Spelling as a Literacy Skill

Spelling is an important literacy skill. Misspellings not only prevent clear communication, but they also have an adverse effect on others' judgment of a person's intelligence (Varnhagen 2000). Some researchers have attempted to address the question of whether spelling has been made irrelevant with the advent of spell checkers, finding that spelling is indeed important (Varnhagen 2000). Spell checkers are flawed, being less effective for misspelled words that are dissimilar to the word orthographically (dissimilar spelling) or phonologically (dissimilar sounding) (Montgomery, Karlan, and

lan, and Coutinho 2001). For example, if someone was attempting to spell "Qatar", a spell checker may not give a correct suggestion if a spelling such as "Katar" was used, as it is spelled differently than the correct word and the sounds are slightly different than the correct word. In contrast to the perspective that technology makes spelling less important, as society becomes more technologically based, spelling increases in importance, especially for online information retrieval (Varnhagen et al., forthcoming). Internet and OPAC searches are included in that online information retrieval.

### *1.2 Spelling and Computer Searching*

Any search performed with a computer requires search terms to be entered letter by letter. While each letter added increases the precision of the retrieval, it also increases the likelihood of an input error – which can have detrimental effects on a search (Yee 1991). Accurate spelling is required in retrieval systems that use keywords (Borgman et al. 1995). Computers, with no reasoning capabilities, will search only for what is entered by the user, so individuals bear the burden of performing accurate and skilled searches. Many OPAC users are familiar with search engines, such as Google, that provide a spell check function. Griffiths and Brophy (2005) cite Google's spell check as one of the reasons the search engine is popular with users. Though many users have begun to expect spell checkers in OPACs (Drabenstott and Weller 1996), many OPACs do not include this feature (Borgman et al. 1995), requiring users to verify their spelling through other means.

### *1.3 Frequency of Misspellings*

The frequency of misspellings in online searches is unknown. Although some researchers have identified misspellings as one of the most common errors made by online catalogue users – of all ages and skill levels (Tenopir 1997), others have rejected misspellings as a serious problem in information retrieval (Drabenstott and Weller 1996). The disagreement may be due to a lack of research. Many research studies have been done on the misspellings that exist in library catalogue records and their effects on information retrieval (Randall 1999), however, there has been little research on user misspellings (Proctor 2002). Part of the difficulty

studying misspellings is that these errors must occur in order to study them; misspellings can be a difficult to capture, as they cannot be manufactured. The results of studies conducted, however, vary widely. Frequency of user misspellings in OPAC searches has been found to vary from 6 to 54% (Yee 1991). Some of this variation, however, may be due to a lack of a clear operational definition of misspellings, as many studies include misspellings, typos and "user errors" which can consist of incorrect spacing or incomplete citations (Yee 1991, 93). It can be difficult to determine whether the spelling is a valid variant spelling, an archaic spelling, a typo or a misspelling. Some of the variation in misspellings might also be due, in part, to the difficulty level of the search term itself. Certain words are more difficult to spell than others (Varnhagen et al., forthcoming). For example, when studying influential 20<sup>th</sup> century thinkers it is easier to search for the works of Sigmund Freud than Michel Foucault. Researchers have found that longer, less familiar and more difficult to spell words result in less successful online searches (e.g. Borgman et al. 1995; Chen 2003).

### *1.4 Impact of Misspellings*

Whatever their prevalence, misspellings can seriously affect information retrieval. Misspellings can reduce recall rates and influence search behaviours, and can even cause the complete failure of a search (Tenopir 1997). While misspellings may cause the failure of a search, users may also cause that failure. Borgman et al. (1995) found that children were more likely to abandon searches that used difficult-to-spell keywords. Griffiths and Brophy (2005) found that even when a good search strategy was used, misspellings reduced the number of results retrieved and users were more likely to discard the search strategy. Part of the difficulty for the user is the lack of results when there is an input error (Sridhar 2004) – a result that often baffles users, who are not sure what to do next to revise the search. Misspellings can also lead to changes in search behaviour. Drabenstott and Weller (1996) observed that informing OPAC users that their searches could contain spelling errors led some users to enter queries on different topics (26%), some to enter the same query (20%), some to quit their search (13%) and some to change or correct their spelling (12%). Misspellings during searches can have great impact on search behaviour.

### 1.5 Online Search Behaviour of Students

Information seeking is dynamic – it depends on both the context and the individual (Heinström 2004). Heinström (2004) discusses the influence of cognitive abilities, cognitive styles, feeling and personalities on information seeking behaviour. While Heinström categorizes users' behaviours and strategies into fast surfers, broad scanners and deep divers, Siegler (1986, 1999) discusses strategy use in general terms more as a choice from many strategies in an individual's repertoire. Siegler (1986) explains how an individual can perform the same task differently at different times. Individuals have different strategies in their repertoire and those strategies differ in the time they require, the demands made on memory, the accuracy of results and the range of problems for which they are appropriate. Because individuals have more than one strategy available, they choose strategies according to the situation – harder problems elicit more overt strategy use and the use of more strategies. Individuals can choose adaptively between strategies – more adaptive strategies are chosen as age and experience increases (Siegler 1999). Spelling is an area to which strategy choices are applied (Siegler 1986).

Searching is not an easy task; it is a skill that must be learned. Searching requires conceptual knowledge of the information retrieval process, semantic knowledge of how to implement the search and technical knowledge about skills and syntax (Borgman 1996). While searching in any venue and with any medium presents challenges, searching for information in the library presents particular problems. After reviewing the literature, Case discusses the problems of intimidation of the search process, deciding the "aboutness" of the search and categorizing similar topics (2007, 22-25). Whether searching on the Internet or in an OPAC, it is the users who are required to shoulder the responsibility of the search, as they are the ones who translate queries into search statements (Borgman 1996).

Students are a group that is required to search; however, they in particular have been identified as having difficulties performing online searches. Griffiths and Brophy (2005) found that 35% of post-secondary students found it difficult to locate information during a search task. One of the reasons for the difficulty experienced by students in their OPAC searching is that they attempt to search the OPAC as they would a search engine, using high levels of keyword searches and low levels of subject headings, Boolean operators and truncation. Student searches are also made difficult as they

searches are also made difficult as they use ineffectual strategies such as changing the search field and adding in keywords (Blecic 1998, 1999; Novotny 2004). The common use of search engines not only changes how users search but also changes their expectations of how a system like an OPAC will work (Novotny 2004). Students are known to make spelling errors during their OPAC searches; however, research also suggests ways that technology can help students identify and correct their spelling (Drabenstott and Weller 1996). Despite this work, problems still remain.

## 2. Current Study

The current study is based, in part, on Varnhagen et al.'s study "Spelling and the Web" (forthcoming). Varnhagen et al.'s study asked university and Grade 4 students to search the Internet using Google. The search topics were given to the students, one of which was easy to spell (lemming) and one of which was difficult (ptarmigan). Ptarmigan was used because Google's "Did you mean:" function would not suggest the correct spelling of the word. Varnhagen et al. found that when unable to spell the search terms, both children and adults changed their search strategies. In particular, adults changed their search strings, leading to more successful searches. Adults and children who had difficulty spelling the search terms viewed more results pages and took more time to complete their search. If participants were unable to spell the search terms, the search was unsuccessful. Few participants used external aids for spelling. This research also discussed previous research and again recommended better spell checking functions.

This study attempted to modify Varnhagen et al.'s study to examine information retrieval using an Online Public Access Catalogue. The goal of the study was to determine what effect misspellings have on catalogue record retrieval when using an OPAC. This study also aimed to examine search behaviours when search terms used during OPAC searches were misspelled. The study consisted of a pilot study and the principal study, both of which have been completed. As only small changes were made to the principal study as a result of the pilot, the results of both phases are examined in this paper.

The current study addresses several questions: How do misspellings in OPAC searches affect information retrieval? How do users' search behav-

iours change when they are unsure of the proper spelling of a search term? What strategies do participants employ when misspelled search terms render OPAC searches unsuccessful? Do the search behaviours of participants searching for difficult-to-spell search terms differ from those searching for easy-to-spell search terms?

### 3. Methods

#### 3.1 Participants

The participants of this study were undergraduate and graduate students at the University of Alberta who were over the age of 18. The scope of this research was limited to university students – a group of people expected to have solid search skills. Participants who were recruited were randomly assigned to the Difficult Spelling or Easy Spelling Group based on the order of the sessions (the participant in the first session was assigned to the Easy Spelling Group, the participant in the second session was assigned to the Difficult Spelling Group, and so on). Each participant was assigned a participant number in order to ensure anonymity. Ethical approval for research involving human subjects was obtained for this study.

#### 3.2 Materials

Participants in the study were given several different tasks: a pre-search checklist, a computer-based search task and a post-search semi-structured interview. The pre-search checklist was given to obtain participants' demographic information, comfort using computers, experience in online searching (including the Internet, OPACs, the University of Alberta's OPAC and databases) and confidence in searching abilities. The search task consisted of a list of 4 search terms. The list of words used in the search task contained two topical, one geographic name and one personal name, according to the frequency of keyword access points used in OPAC searches (Drabenstott and Weller 1996). There were two versions of the lists, one for the Easy Spelling Group and one for the Difficult Spelling Group. The Easy Spelling Group received words at approximately a Grade 10 reading level. These words included lemming, civilian, Bolivia and Sigmund Freud. The Difficult Spelling Group received words from published lists of commonly misspelled words and, in the case of the geographical and personal names, words that are orthographically impossible in English. These words included

ptarmigan, millennium, Qatar and Michel Foucault. Lemming and ptarmigan were chosen because they were the search terms from Varnhagen et al.'s study (forthcoming). The checklist and search tasks were created using research tools from the literature (Slone 2000; Varnhagen et al., forthcoming) and new questions designed specifically for this study.

A semi-structured interview was used to explore participants' confidence levels when searching, any troubles they encountered during the search task, any difficulties spelling search terms and the reasons behind the participants' search behaviours. The semi-structured interview was created using the aforementioned research tools from the literature, questions previously used for a practice interview and new questions designed for this study.

#### 3.3 Procedure

Participants began by filling out the pre-search checklist (Slone 2000). Then, participants were given up to 20 minutes to perform a search task in the University of Alberta's OPAC, which is run by the SirsiDynix software Unicorn and which is a union catalogue from Networking Edmonton Online System (NEOS) – a consortium of central Alberta government, health, college and university libraries. The OPAC was accessed through the participant's web browser of choice, either Internet Explorer 7.0 or Firefox 2.0.

The researcher read the first search term to the participants. The researcher then asked participants to find as many items in the catalogue for the topic as they were able. If participants did not know what the topic was, they were given a brief definition (e.g. "A ptarmigan is an arctic bird"). Once participants indicated that they had finished the search, the researcher gave them the next search term. The computer used for the search was equipped with the screen recorder program Camtasia to capture all the activity on the computer screen and whatever comments the participants made while searching (Novotny 2004; Varnhagen et al., forthcoming). After the search task, participants were asked to complete a post-search semi-structured interview (Slone 2000). The interviews were also recorded using Camtasia.

### 4. Results and Discussion

#### 4.1 Participant Demographics

Information about the participants was gathered through the pre-search checklist. The checklist

asked for demographic information. There were 38 students who participated in the study, 6 in the pilot and 32 in the principal study, including 24 (63%) females and 14 (37%) males. The median age was 19.5 (range 18-59). The participants included 6 graduate students and 32 undergraduate students and had a variety of years of school experience. Undergraduate and graduate students ranged from 1 to 4 years in their current programs. Participants were divided into junior students (students in their first and second years of university) and senior students (students in their third year of university or higher), of which there were 20 (53%) junior students and 17 (45%) senior students (one student did not provide that information). The students had a variety of backgrounds, with the majority of students coming from the Faculty of Science (17), followed by Arts (7), Engineering (5), Education (4), Physical Education (4) and Open Studies (1).

In addition to the demographic information, the pre-search checklist asked participants about their comfort level with computers, the frequency of their searching and their confidence in finding resources for which they are looking during searches. In general the participants were comfortable with computers. Participants rated themselves on a 4-point Likert scale that ranged from "Very Comfortable" (a rating of 4) to "Not Comfortable" (a rating of 1). The average rating was 3.5, as 19 participants (50%) rated themselves as "Very Comfortable" and the other 19 participants (50%) rated themselves as "Comfortable." Participants in the Difficult Group rated themselves as slightly more confident than those in the Easy Group.

Participants rated the frequency of their searching on a 5-point Likert scale from "Every day" (a rating of 5) to "Never" (a rating of 1). Participants most frequently searched the Internet, averaging a rating of 4.4, less frequently searched the University of Alberta OPAC (averaging a rating of 2.2), less frequently still the U of A's databases (averaging a rating of 2.1) and almost never searched other libraries' OPACs (averaging a rating of 2.0). While the two groups were almost identical in frequency of use, the Difficult Group reported their use of the Internet as more frequent than the Easy Group.

While comparing participants by demographic information, comfort using computers, experience in online searching and confidence in searching abilities created groups that were too small to determine any statistically significant differences, trends were examined. Any participant differences between groups were very small. Overall participants were comfortable using computers and had some experience doing online searching.

#### 4.2 Internet Searching

Participants were asked to perform 4 separate search tasks to find as many items in the University of Alberta library catalogue as they could for given search topics. All participants were told they could use any Internet resource to aid them in their OPAC search if they desired, therefore, the search tasks given to participants were performed on both the Internet and the OPAC. Participants used several different Internet and computer resources to check their spelling, to ensure they were using the correct terms and to gather background information about the topics. (Often, participants were more than simply checking to see if they had the correct spelling, they were trying to find the term, as they were completely unfamiliar with it.) There were differences between how the two groups used the Internet. For each search, the Difficult Group checked spelling more often. Of the 19 participants in the Difficult Group 18 (95%) used the Internet, compared to 8 (42%) of the Easy Group. Less than half of the participants in the Easy Group searched the Internet. While participants did not check the spelling of all the search terms, participants checked the spelling for a total of 72 searches (47% of the 152 searches). (For the breakdown of each search performed on the Internet to check spelling and gather background information, please see Table 1.) There was no discernable difference between the groups for checking background information.

Interestingly, there was a difference between junior students (students in their first and second years of university) and senior students (students in their third year of university or higher). Junior students checked the spelling on 35% of their searches (28 of the 80 searches). Senior students checked

Table 1. Number of participants per spelling group who searched the Internet to check spelling and gather background information.

Search Term	Checked Spelling (%)		Background Information (%)	
	Easy	Difficult	Easy	Difficult
1. Lemming / Ptarmigan	4 (21%)	14 (74%)	3 (16%)	6 (32%)
2. Civilian / Millennium	1 (5%)	10 (53%)	3 (16%)	3 (16%)
3. Bolivia / Qatar	3 (16%)	18 (95%)	4 (21%)	4 (21%)
4. Sigmund Freud / Michel Foucault	4 (21%)	18 (95%)	2 (11%)	2 (11%)

**Proceedings of the 36th annual conference of the Canadian Association for Information Science (CAIS), University of British Columbia, Vancouver, June 5-7, 2008**

the spelling on 60% of their searches (41 of the 76 searches). The difference was also seen in the students who performed searches for background information. Junior students gathered background information on 9% of their searches (7 of the 80 searches). Senior students gathered background information on 26% of their searches (20 of the 76 searches). Years of education affected search behaviours as senior students performed more spelling checks and background information searches than junior students.

Students who searched the Internet more often also tended to check the spelling of their search terms. Of 60 searches performed by the 15 students who search the Internet between 2 and 5 times per week, 23 (38%) included a spell check. Of the 76 searches performed by the 19 students who search the Internet every day, 45 (59%) included a spell check. Experience using the Internet also appears to affect search behaviours, however this finding could also be attributed to group differences as the Difficult Group (who checked spelling more often) also contained slightly more participants who searched the Internet more frequently. This finding should be further explored to determine whether experience searching increases awareness that correct spelling influences information retrieval.

#### 4.3 Internet Searching Strategies

Participants used a variety of search strategies on the Internet in order find the spelling of the term, including: typing in the word, trying spelling variations, typing in the word and the topic together, and typing in the topic by itself. (For the breakdown of the search strategies performed on the Internet to check spelling, please see Table 2.) Participants

who had difficulty finding the spelling of the search term initially, often changed search strategies and searched for the general subject, rather than the specific noun. For example, "Middle East" was used in Wikipedia to find "Qatar" and "French philosopher" was used in Google to find "Michel Foucault." Spelling variations of the term were also paired with the subject in order to find the spelling. For example, "tarmagin, bird" was used in Google to find "ptarmigan."

Only 1 participant in the Easy Group used the general topic (either by itself or in combination with a variation of the word) to find the spelling of the search term. Using the topic as a search strategy to find the spelling was used in 28 (37%) of the Difficult Group's searches (including using the word and the topic, the topic by itself or a topic search as one of multiple search strategies). All other search strategies used by the Easy Group to check the spelling were typing in the word itself (in some variation). While participants in the Easy Group more often simply typed in the word, those in the Difficult Group used multiple strategies – in particular including the search topic as part of the search query. The inclusion of search topics to find difficult-to-spell words was observed in other research (Varnhagen et al., forthcoming).

There was also a difference between junior students' and senior students' search strategies, with the senior students tending to use more topic searches. (For the breakdown of the search strategies performed on the Internet by junior and senior students to check spelling information, please see Table 3.) Using the topic as a search strategy to find the spelling was used in 19 (28%) of the senior students' searches compared to 9 (11%) of junior

Table 2. Number of participants per spelling group who used different search strategies to check spelling.

	Type in Word Only	Try Spelling Variations	Type in Word and Search Topic Together	Type in Search Topic	Use Multiple Strategies	Total
<b>Easy</b>	11 (92%)	0 (0%)	1 (8%)	0 (0%)	0 (0%)	12 (100%)
<b>Difficult</b>	25 (42%)	6 (10%)	11 (18%)	3 (5%)	15 (25%)	60 (100%)

Table 3. Number of junior and senior students who used different search strategies to check spelling.

	Type in Word Only	Try Spelling Variations	Type in Word and Search Topic Together	Type in Search Topic	Use Multiple Strategies	Total
<b>Junior</b>	14 (50%)	5 (18%)	1 (4%)	1 (4%)	7 (25%)	28 (100%)
<b>Senior</b>	21 (49%)	1 (2%)	10 (23%)	2 (5%)	9 (21%)	43 (100%)

students. Inclusion of the topic in the search appears to be affected by the term difficulty as well as the years of education of the searcher.

Participants used a combination of checking the spelling and retrieving the spelling from memory to obtain the correct spelling. Overall the strategies used were effective. Of the 152 searches, 123 (81%) were found either by retrieval from memory or by looking the search term up on the Internet. Some of the correct spellings that were obtained were a result from correcting initially misspelled search terms. Of the 123 that were found, 19 (15%) were corrections made after attempting to search using misspellings. The majority of the corrections made were as a result of checking spelling (16 corrections or 84%), while the rest (3 corrections or 16%) were made from memory. Of the searches that contained a misspelling (millennium), 60% were not caught because the spelling was not checked. (Checking spelling did not always result in correct spelling, however as the spelling was checked for 40% of the searches that contained the misspelled "millenium". Misspellings of millennium, however, were not caught by all spell checkers and were even offered as a suggested spelling, attesting to millennium as a commonly misspelled word.) Despite the problems obtaining the correct spelling of millennium, the search strategies of checking spelling or retrieving correct spelling from memory were effective.

#### *4.4 Resources Used to Check Spelling*

Participants used a variety of sources to check spelling. The participants used some of the more common Internet resources such as Google, Wikipedia and Dictionary.com. Of the 23 participants who used other resources to check their spelling, 21 used Google, 5 used Wikipedia, 4 used Dictionary.com, and 1 used Merriam-Webster, Yahoo!, Ask.com and Microsoft Word. (It should be noted that Google was the search tool located on the menu bar of both the web browsers used for the study – Internet Explorer 7.0 and Firefox 2.0). Of the 13 participants who used other resources to get background information on the topic, 9 used Wikipedia, 5 used Google, 3 used websites found through Google and 1 used Dictionary.com. The resources used by the different groups were similar, however it is difficult to compare as more participants in the Difficult Group performed Internet searches than participants in the Easy Group. Participants used the different information sources in different ways. Wikipedia was used to ensure an

unfamiliar term was the one for which they were looking, as well as to gather background information. Google was used to search for the term, as well as for its spelling suggestions. While some of the most common websites were used, they were used in more sophisticated ways, to meet specific needs for the task at hand.

The spelling suggestions were obtained through Google's automatic "Did you mean:" function or through typing a word into the Google search box on the menu bar (as you input letters, a list of possible words is given in a drop down menu below). Google automatically "checks to see if you are using the most common version of a word's spelling" (Google, 2007, para. 22). The "Did you mean:" function appears when Google calculates that a more relevant search would be generated using an alternate spelling. Google's suggestion is based on all the occurrences of the words used in the Internet. Similar "Did you mean" functions could be obtained with Dictionary.com and Merriam-Webster. Of the 152 searches performed, 36 searches (24%) used spelling suggestions. Of the 36 spelling suggestions, 31 (86%) were from Google (either "Did you mean" or the search box suggestions), 4 (11%) were from Dictionary.com and 1 (3%) was from Merriam-Webster. Interestingly, Google did not suggest "millennium" when "millenium" was used as a search term, indicating this is a common spelling mistake. Participants used the spelling suggestions generated by Google and Dictionary.com; however the spelling suggestions were not always correct.

#### *4.5 OPAC Searching*

Each participant had 4 search tasks to complete using the University of Alberta's OPAC. In total, there were 152 searches made, 138 (91%) of which were completed. The searches not completed were due to the participant not being able to retrieve information on the topic given. Of the 76 searches performed by the Easy Group, 75 (99%) were completed. Of the 76 searches performed by the Difficult Group, 63 (83%) were completed. Of those 152 searches, 46 searches used misspelled search terms at some point during the OPAC search, meaning 30% of searches had a misspelled search term. (For the breakdown of each search performed that included a misspelling, please see Table 4.) The searches were categorized as containing a corrected spelling mistake, an uncorrected spelling mistake or being unsuccessful (the search was not completed and information on the topic could not be retrieved).

Table 4. Number of participants per spelling group with misspellings that were corrected, left uncorrected or led to unsuccessful searches.

Group	Misspelling	Words Misspelled (Count)	Total	% of Misspellings	% of Searches
Easy	Corrected	<ul style="list-style-type: none"> <li>• Lemming (2)</li> <li>• Bolivia (2)</li> </ul>	4	9%	3%
	Uncorrected	<ul style="list-style-type: none"> <li>• NA</li> </ul>	0	0%	0%
	Unsuccessful	<ul style="list-style-type: none"> <li>• Lemming (1)</li> </ul>	1	2%	<1%
<b>Total</b>			5	11%	3%
Difficult	Corrected	<ul style="list-style-type: none"> <li>• Ptarmigan (5)</li> <li>• Millennium (3)</li> <li>• Qatar (3)</li> <li>• Michel Foucault (2)</li> </ul>	13	28%	9%
	Uncorrected	<ul style="list-style-type: none"> <li>• Millennium (15)</li> </ul>	15	33%	10%
	Unsuccessful	<ul style="list-style-type: none"> <li>• Ptarmigan (4)</li> <li>• Qatar (6)</li> <li>• Michel Foucault (3)</li> </ul>	13	28%	9%
<b>Total</b>			41	89%	28%
<b>Total</b>			<b>46</b>	<b>100%</b>	<b>30%</b>

The searches that contained misspelled search terms retrieved no records, with the exception of “millenium”. The misspelled “millenium” retrieved results from the University of Alberta OPAC as the OPAC contains that spelling in the records. Of the 19 searches for millennium, 15 (79%) were misspelled.

One factor that affected participants’ search success was their familiarity with the given topics. All searches that retrieved no results were searches for unfamiliar items. Participants could not get the correct spelling of these terms. The only familiar topic that resulted in unsuccessful searches was millenium – all participants that misspelled that search term were familiar with the topic. The incorrect spellings of millennium (“millenium”) resulted in decreased information retrieval.

While it was not possible to determine how spelling affected information retrieval of groups (as difference in retrieval rates was due more to search field used rather than search topic difficulty), it was very clear that misspellings greatly affect searches. Misspelled search terms retrieve a fraction of the total number of items on a topic and are only successful in retrieving those few items due to errors in the catalogue (or in the original works). These instances of retrieving records with misspellings often misleads students into believing that they have

found all that they can on the topic. This appeared to be the case for 79% of the participants who completed a search for the topic of millennium with a misspelling. For those misspellings that do not occur in the OPAC, no results are retrieved. If the correct spelling cannot be obtained in some way, as was the case with 9% of the total searches, then the search results in a failure to retrieve any information on the specific topic, confirming the findings of other studies (e.g. Drabenstott and Weller 1996; Varnhagen et al., forthcoming). More difficult-to-spell and unfamiliar terms result in misspelled search terms and searches that retrieve no results.

#### 4.6 OPAC Search Strategies

Participants’ catalogue searching was examined to determine the search strategies used to find the topics. Several different search strategies were examined including using: spelling variations, Boolean operators, truncation, synonyms, other search fields, additional terms, multiple searches and the same searches retyped. (For the breakdown of the search strategies performed using the OPAC to find items on the topic, please see Table 5.) While some of these strategies were used at the beginning of the searches with the intent to find the spelling, only 2 searches were able to find resources specifically on the topic using general searches or



spelling variations. The majority of strategies employed were used to find items on the topic. The most common search strategies were adding in search terms, performing multiple searches using the same search term with other search fields. Search strategies were quite similar between the two groups.

The search strategies used by participants are similar to the findings by Novotny (2004). Students used Boolean operators, truncation and synonyms infrequently, while they changed search fields and added terms frequently. As with the students in Novotny's study, this indicates that students may be attempting to search the library OPAC as they would a search engine.

Participants tended to keep their OPAC searches simple. The majority of participants' completed searches consisted of one word (or one full name) in their search. Of the completed 138 searches, 101 searches (73%) contained one search term, only 37 searches (27%) combined two different terms in the search string. The terms added were either an attribute of the topic (e.g. "bird" added to ptarmigan) or a specific aspect of the topic (e.g. "civilian military").

While there was some variety in the search field used, the majority of participants used "Any field" in which to perform their search. (For the breakdown of the search fields used in the OPAC searches to

find items on the topic, please see Table 6.) Participants also tended to keep the search field broad, most often choosing "Any field" as the field in which to search. Of the 138 final searches, 79 searches (79%) used "Any field". The groups used fairly similar search fields, with the exception of the Subject field, which was used more by the Easy Group. The use of search field also did not appear to be affected by a participant's familiarity with the search topic.

Participants who used the Internet more tended to use different search fields than those who searched the Internet less often. There were 15 students who rated their use of the Internet as between 2 and 5 times per week (for a total of 60 searches). There were 19 students who rated their use of the Internet as everyday (for a total of 76 searches). (For the breakdown of the search fields used in the OPAC searches to find items on the topic by frequency of use of the Internet, please see Table 7.) Students who searched the Internet everyday were more likely to use "Any field". Students who searched the Internet between 2 and 5 times per week were more likely to use the "Subject field" or subject headings, or to use multiple fields. The finding that students who use the Internet more often use "Any field" more frequently also indicates that students' OPAC searches are influenced by their experience with search engines.

Table 5. Number of participants per spelling group who used different search strategies to find topics in the OPAC.

Search Strategy	Easy Group Total (% of Group's Searches )	Difficult Group Total (% of Group's Searches)	Total (% of Total Searches)
Spelling Variations	1 (1%)	7 (9%)	8 (5%)
Boolean Operators	8 (11%)	9 (12%)	17 (11%)
Truncation	2 (3%)	0 (0%)	2 (1%)
Synonyms	3 (4%)	4 (5%)	7 (5%)
Change Search Fields	29 (38%)	23 (30%)	52 (34%)
Additional Terms	30 (39%)	39 (51%)	69 (45%)
Multiple Searches	36 (47%)	47 (62%)	83 (55%)
Searches Retyped	9 (12%)	6 (8%)	15 (10%)

Table 6. Number of participants per spelling group who used different search fields to find topics in the OPAC.

Search Field	Easy Group (% of Group's Completed Searches)	Difficult Group (% of Group's Completed Searches)	Total (% of Total Completed Searches)
Any	39 (52%)	40 (63%)	79 (57%)
Author	1 (1%)	0 (0%)	1 (<1%)
Multiple	8 (11%)	9 (14%)	17 (12%)
Subject	20 (27%)	5 (8%)	25 (18%)
Subject Heading	0 (0%)	4 (6%)	4 (3%)
Title	7 (9%)	5 (8%)	12 (9%)

Proceedings of the 36th annual conference of the Canadian Association for Information Science (CAIS), University of British Columbia, Vancouver, June 5-7, 2008

Table 7. Number of participants by frequency of Internet use who used different search fields to find topics in the OPAC.

Group	Any Field	Subject Field and Subject Headings	Multiple Fields
Search Internet 2-5 Times Per Week	21 (35%)	17 (28%)	10 (17%)
Search Internet Every day	48 (63%)	11 (14%)	5 (7%)

Few participants rated their catalogue use as frequent. The majority of participants (22 or 58%) rated their catalogue use as “every few weeks”. A total of 27 participants (71%) rate their use between “never” and “every few weeks”. While it was difficult to examine the results of participants’ searches by their experience using the OPAC, participants with greater experience tended to be able to complete searches more often (i.e. fewer searches were performed that resulted in no records retrieved). The number of searches that contained a spelling mistake did not appear to decrease with experience.

While many factors influence search behaviours and retrieval results, spelling is an important aspect of online searching whether on the Internet or in an OPAC. With the exception of misspelled search terms that retrieved misspellings from the catalogue records, an uncorrected misspelling led to the retrieval of no information on the topic.

## 5. Limitations

Participants were asked to search for a specific search term in an unfamiliar setting and given approximately 5 minutes in which to do so. Participants were both observed and recorded. Because this is not a natural setting in which students perform OPAC searches, the validity of the study is decreased. While university students may perform searches for terms with which they are not familiar, they rarely perform searches with such strict time limits nor use search terms they have not seen. (However, there have been cases of students searching for information heard in class with very poor outcomes (Dewdney and Michell 1996).)

The fact that the participants were observed may cause participants to behave in ways that are not typical. Participants were encouraged to perform searches as they normally would and were made to feel as comfortable as possible during the study session. However, some participants expressed that they rarely or never search the library catalogue, their searches are restricted to journals and databases, so searching the catalogue could not be

done in a “normal” way. Also the instructions participants received were altered slightly during the course of the study as some participants were unsure of how to indicate their selected items from the OPAC search. While the instructions given were then clearer, the modified instructions may have altered participants’ behaviours.

The factors mentioned above may interfere with studying how misspellings during OPAC searches affect information retrieval and search behaviour. Because the results obtained may be influenced by the lab environment, the study’s overall generalizability is limited.

## 6. Implications

### 6.1 Spell Checkers in OPACs

The call for spell checkers in OPACs is not new (Drabenstott and Well 1996), however it is a call that should be repeated. When you compare the number of items currently retrieved by an “Any field” search for *millennium* (2496) and for *millenium* (152), the damaging impact of misspellings on information retrieval cannot be ignored. While searchers who are more experienced with particular topics might recognize that the number of results for a search is incorrect, leading them to examine their search terms, less experienced searchers may not. And while misspellings may drastically reduce the number of items retrieved by a search, what occurs more commonly is the recall of no items. Misspellings only recall items that have been misspelled in the catalogue. While the number of items currently retrieved by an “Any field” search for *Michel Foucault* is 498, a search for *Michelle Foco* (a common misspelling for those not familiar with the French philosopher) will retrieve 0 records (and the list of suggestions will not be of help as the suggestions are arranged alphabetically by last name).

Because of the serious effects of misspellings on information retrieval, librarians need to advocate for the inclusion of spell checkers with vendors. While

spell checkers have been common in word processors for decades and are now standard in search engines (where most students get their online search experience), most OPACs do not have spell checkers. The technology exists for spell checkers in integrated library systems, however most systems currently do not include spell checkers. SirsiDynix, one of the most common vendors of Integrated Library Systems (ILS) to academic libraries is not adding spell checking as a new feature in their new ILS software Symphony. When choosing library software, librarians must be aware of the importance of spell checkers and to advocate for their inclusion in systems that do not yet have them.

### 6.2 Cataloguing

While most errors in catalogues do not prevent access to records (Randall, 1999), those errors can be retrieved by misspelled search queries. As advocated by Ballard and Lifshin (1992), libraries need to systematically review their records to determine whether they contain spelling mistakes. There are additional considerations, however, such as the inclusion of records with non-English words and archaic spellings (Ballard and Lifshin 1992). While the search for "millenium" currently retrieves 152 records, some of those records are misspelled, some refer to archaic spellings and others are misspellings that are noted with [sic]. These variations make a complex situation and raises issues of how to indicate to library users why the spelling variations occur. However, the situation can be simplified to a certain degree by fixing errors to maintain authority control. Ballard and Lifshin (1992) also point out that part of the solution is raising the awareness of cataloguers.

### 6.3 Information Literacy Instruction

Librarians doing information literacy instruction (ILI) have many aspects of student searching to address, such as students searching OPACs the same way in which they search the Internet. While there are many important issues vying for attention, knowing that currently the majority of OPACs do not have spell checkers and that OPACs are imperfect (and will contain spelling mistakes), spelling needs to be an issue brought to students' attention. Students can begin to feel overwhelmed by the complexities of OPAC searching and librarians therefore

make attempts to limit the information provided, only to essentials, but spelling simply needs to be on the table. Performance Indicator 2 of Standard Two of the *Information Literacy Competency Standards for Higher Education* states that "The information literate student constructs and implements effectively-designed search strategies," and goes on to state that that includes outcomes such as identifying synonyms and using controlled vocabulary (ALA 2000). As many librarians include issues around search strategies, this could be the point at which the issue of spelling could be bridged. Spelling is only one issue among many, however, when it is ignored it can greatly affect information retrieval.

Spelling is an important topic that is often ignored. While many students do not think about spelling when searching online, misspelling search topics can have serious effects on information retrieval. Spelling needs to be considered when choosing an OPAC, inputting catalogue records and teaching information literacy skills to students. While students have a number of strategies in their repertoire and can use them adaptively, misspellings can still adversely affect information retrieval.

## 7. Future Directions

In this paper students' searching behaviours were examined. Semi-structured interviews were also conducted after the search tasks. In the future, the interview data will be examined in more detail – in particular, participants' normal search behaviour, problems (including spelling) encountered while searching, the frequency with which participants search for relatively unfamiliar topics they have only heard about and not seen, confidence that all relevant information was retrieved, the reasons behind search behaviours and the utility of a spell check in the University of Alberta's OPAC.

While the analysis of this study will continue, the possibility of examining transaction records for searches performed at the University of Alberta would allow some unobtrusive observations of students' search behaviours and their handling of misspellings to confirm or disconfirm the results from obtrusive observations. While transaction records have been studied widely, few have studied students' handling of misspellings.

## Acknowledgements

I would like to thank the Social Sciences and Humanities Research Council of Canada and the Government of Alberta for funding this project. I would also like to thank the faculty and staff of the School of Library and Information Studies at the University of Alberta for supporting this research. RJW

## References and bibliography

- American Library Association. Information Literacy Competency Standards for Higher Education. Association for College and Research Libraries. <http://www.ala.org/ala/acrl/acrlstandards/informationliteracycompetency.cfm>.
- Ballard, T., and A. Lifshin. 1992. Predictions of OPAC spelling errors through a keyword inventory. *Information Technology and Libraries* 11, no. 2: 139-45.
- Blecic, D. D., J. L. Dorsch, M. H. Koenig, and N. S. Bangalore. 1998. A longitudinal study of the effects of OPAC screen changes on searching behavior and searcher success. *College & Research Libraries* 60, no. 6: 515-30.
- Blecic, D. D., N. S. Bangalore, J. L. Dorsch, C. L. Henderson, M. H. Koenig, and A. C. Weller. 1998. Using transaction log analysis to improve OPAC retrieval results. *College & Research Libraries* 59, no. 1: 39-50.
- Borgman, C. L. 1996. Why are online catalogs still hard to use? *Journal of the American Society for Information Science* 47, no. 7: 493-503.
- Borgman, C. L., S. G. Hirsch, V. A. Walter, and A. L. Gallagher. 1995. Children's searching behavior on browsing and keyword online catalogues: The science library catalogue project. *Journal of the American Society for Information Science* 46, no. 9: 663-84.
- Chen, S. L. 2003. Searching the online catalogue and the World Wide Web. *Journal of Educational Media and Library Sciences* 41, no. 1: 29-43.
- Dewdney, P., and G. Michell. 1996. Oranges and peaches: Understanding communication accidents in the reference interview. *Reference Quarterly* 35, no. 4: 520-34.
- Drabenstott, K. M., and M. S. Weller. 1996. Handling spelling errors in online catalogue searches." *Library Resources and Technical Services* 40, no. 2: 113-32.
- Heinström, J. 2005. Fast surfing, broad scanning and deep diving: The influence of personality and study approach on students' information-seeking behavior. *Journal of Documentation* 61, no. 2: 228-47.
- Google. Spell checker. <http://www.google.com/help/features.html#spell>.
- Griffiths, J. R., and P. Brophy. 2005. Student searching behavior and the web: Use of academic resources and Google. *Library Trends* 53, no. 4: 539-54.
- Malliari, A., and D. Kyriaki-Manessi. Users' behaviour patterns in academic libraries' OPACs: A multivariate statistical analysis. *New Library World* 108, no.3/4: 107-22.
- Montgomery, D. J., G. R. Karlan, and M. Coutinho. 2001. The effectiveness of word processor spell checker programs to produce target words for misspellings generated by students with learning disabilities. *Journal of Special Education Technology* 16, no. 2: 27-41.
- Novotny, E. 2004. I don't think I click: A protocol analysis study of use of a library online catalog in the Internet age. *College & Research Libraries* 65, no. 6: 525-37.
- Proctor, E. 2002. Spelling and searching the Internet: An overlooked problem. *The Journal of Academic Librarianship* 28, no. 5: 297-305.
- Randall, B. N. 1999. Spelling errors in the database: Shadow or substance? *Library Resources and Technical Services* 43, no. 3: 161-69.
- Siegler, R. S. 1986. Unities across domains in children's strategy choices. In *Perspectives on Intellectual Development: The Minnesota Symposium on Child Psychology*, ed. M. Perlmutter, 1-48. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Siegler, R. S. 1999. Strategic development. *Trends in Cognitive Sciences* 3, no. 11: 430-35.
- Slone, D. J. 2000. Encounters with the OPAC: Online searching in public libraries. *Journal of the American Society for Information Science*, 51(8), 757-73.
- Sridhar, M. S. 2004. OPAC vs card catalogue: A comparative study of user behaviour. *The Electronic Library* 22, no. 2: 175-83.
- Tenopir, C. 1997. Common end user errors. *Library Journal* 122:31-32.
- Varnhagen, C. K. 2000. Shoot the messenger and disregard the message? Children's attitudes toward spelling. *Reading Psychology* 21:115-28.
- Varnhagen, C. K., G. P. McFall, L. Figueredo, B. S. Takach, J. Daniels, and H. Cuthbertson. Forth-

coming. Spelling and the web. *Journal of Applied Developmental Psychology*.  
Yee, M. M. 1991. System design and cataloging meet the user: User interfaces to online public access catalogues." *Journal of the American Society for Information Science* 42, no. 2: 78-98.