Initial Articles in Library Catalog Title Searches: An Impediment to Information Retrieval

Edward M. Corrado
TCNJ Library, The College of New Jersey, PO Box 7718, Ewing, New Jersey, 08628-0718, USA.  corrado@tcnj.edu

Initial articles have posed difficulty for end-users since the first online library catalogs. This study examines the way users search using left-anchored title searches for items that contain insignificant initial articles in a library catalog. It also investigates how users react when they do not find an item because their search phrase included an insignificant initial article. The results show that a considerable number of users do not find existing titles that have insignificant initial articles because their search strategy does not coincide with the way the Library Management System was designed to handle index and retrieve these titles. Suggestions for further research are offered.

Introduction

Library cataloging rules inform librarians and library system designers that insignificant initial articles should be ignored when filing, indexing, or searching for titles. There are various ways library management systems (LMS) deal with title searches that include initial articles. For the purposes of this study, title searches are defined as left-anchored searches of indexes that consist of entries that include the title proper, variant titles, and series titles. Many systems will perform the search as entered, initial articles included, which would lead to a failed search since the initial articles are not indexed. In some cases the searcher may not know when a word that is often treated as an insignificant initial article should or should not be considered significant. For example, "Los" is an article and normally should be dropped from a title search if it is the first word, however if it is used as the first part of the place name Los Alamos, such as in the title Los Alamos Wildfires by Nichol Bryn, it should not be dropped. While this may not cause a problem for more experienced searchers, many people who search a library catalog are perpetual novices (Borgman, 1986).
In order to assist patrons, many libraries have added text to the title search name or additional directions to inform patrons what to do when searching for titles with initial articles. Some libraries, for example, will label their title search to something resembling "Title (omit initial articles)." Research has shown, however, that many users do not understand library jargon such as this.¹ Do users actually omit initial articles while searching? Furthermore, if users do a title search that includes an initial article and do not find what they are looking for, do they go back and search again without the initial article? This study analyzed the Fall 2005 semester search logs from one American mid-sized academic library catalog to help determine the answers to these questions. One hundred twenty title searches that included an initial article were thoroughly examined to see if, 1) it was a proper use of an initial article, 2) if the user modified the search to search without the initial article, and 3) in the cases where the user didn't modify the search, would a modified search have yielded the correct hit. The library whose logs were analyzed uses the Voyager LMS from Endeavor Information Systems.

Background

The impetus for this investigation was an inquiry from a member of the teaching faculty about how the LMS handled title searches that contained initial articles, and if it could be adjusted to become more user-friendly. Librarians wondered if this might be an isolated incident or if it was a larger issue. A search of the library literature revealed little research to base a recommendation on so it was decided that a review of the LMS search history should be conducted to see if this was an isolated incident or not.

The practice of ignoring initial articles in library catalogs goes back to the days before computers. In library card catalogs insignificant initial articles were not indexed, in part, because the sheer volume of titles beginning with them would clutter up the catalog and make it unwieldy to find the title in question. This practice was then continued in the days of computer generated microfiche and print catalogs and continues in today's computerized library catalogs.

Some library catalogs search for the title as typed, including the initial article, even though the title will not be indexed with the initial article. This flawed search strategy results in a failed search even if the catalog includes the title. Other implementations of systems will automatically remove any initial articles a user inputs into a title search. If a catalog drops all initial words that are normally considered insignificant articles, title searches that begin with a word that in a particular case is not an insignificant initial article, will fail. For example, if the catalog automatically deleted the word "A" in a search for the title "A" is
for Alibi, by Sue Grafton the search would became is for alibi and would not locate the book in the catalog. This can also cause problems when an acronym is spelled the same way as an initial article (Arsenault & Ménard, 2005). Both of these, and methods used by other systems, have drawbacks. Also, because different library catalogs employ different methods, confusion may arise for people who are familiar with one catalog when they switch to another library's catalog, and can have a negative effect when searching different catalogs simultaneously.

Nielsen and Pyle (1995) wrote that although “Initial articles have caused problems for librarians and library patrons for many years” yet there has been “next to nothing about this perennial problem” in the library literature. Despite their hope that their study would foster more research in this area, little research besides Arsenault and Ménard's 2005 study have been conducted. This is disappointing because searching by title is a very common, and important, method for accessing library collections and issues caused by initial articles have not disappeared. This study will hopefully show that not only has the problem not gone away, but that the problem may be bigger than many in the profession realize.

Methodology

Title searches, recorded in web server logs, performed between September 1, 2005 and December 22, 2005 that started with any of the three common initial articles in English (a, an, the) where chosen at random to be analyzed. Because the library catalog contains mostly English titles and English is the native language of most of the users, only initial articles in English were analyzed in hopes to limit the effect of the users' knowledge, or lack thereof, of foreign languages on search strategies. A number was assigned to each title search that began with an initial article and a random number generator picked 128 of these searches to be analyzed 2. Eight of these searches were discarded because they either appeared to be non-title searches or contained gibberish resulting in 120 searches being included in the analysis.

The LMS assigns each search session a unique process identifier (PID) which is recorded in the log. The PID was used to determine a search session. The searches in each session before and after the randomly chosen search were investigated. Other searches from the same time period were also examined to determine if it appeared the user may have tried to use another computer or get help from a reference librarian (or another library user). In particular, attempts by the user to modify the title search or to use a different type of search (an author or keyword search, for example) to locate the title were looked for. Searches performed by the user were duplicated during analysis to see if the user found
the item they were looking for in the library catalog. After investigating to see if the search was or was not modified, the catalog was searched to determine if the title exists in the catalog and whether a properly modified search would have found the title.

Results

A total of 276,620 searches were performed during the time period analyzed. Of these, 99,838 were title searches, with 3,660 of the title searches beginning with an initial article in English (see Table 1).

<table>
<thead>
<tr>
<th></th>
<th>September</th>
<th>October</th>
<th>November</th>
<th>December</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Title searches</strong></td>
<td>24469</td>
<td>28141</td>
<td>28611</td>
<td>18617</td>
<td>99838</td>
</tr>
<tr>
<td><strong>Title searches containing initial articles</strong></td>
<td>1132</td>
<td>1011</td>
<td>977</td>
<td>540</td>
<td>3660</td>
</tr>
</tbody>
</table>

In only 42.5% of the title searches analyzed did the user attempt to modify the search (see Figure 1). When users correctly modified a search, they found the title in the library catalog 80% of the time. When a search was not modified, the library owned the title approximately 62% of the time - however the user appears never to have found the item in the catalog (see Figure 2). Overall, it was found that when a user started looking for a title (that the library owned) with a search that included an initial article, the failure rate was over 51%.

![Image of pie chart showing 42.5% User Did Not Modify Search and 57.5% User Modified Search]

Figure 1: Percentage of failed searches beginning with Initial Articles that were modified (n=120)
Figure 2: Status of library holdings when search was not modified (n=89)

An analysis of bibliographic records in the library catalog revealed that over 22% of the titles in the catalog start with an initial article. If one assumes that users search for titles containing initial articles at the same rate as they exist in the catalog, almost 18% of the time these titles are searched for the initial article is included in the search. Since the failure rate of these searches when the first search included an initial article is just over 51% this means that approximately 9% of the time users did not find a title that was in the catalog that begins with an initial article because of improper searching techniques for the LMS software in use.

The log files showed that users tried different strategies to modify a search when they failed to find an item when they included an insignificant initial article. In the cases where they modified the search in a proper manner, the most common method was to perform the search without the initial article. This method was employed approximately 60% of the time. Other common methods included searching by author or using a keyword search.

In other cases, users tried unsuccessful strategies while attempting to modify a title search. One strategy used on more then one occasion was to change the capitalization of the words in the search. For example, if a user performed a title search for THE TRUTH
ABOUT CATS AND DOGS with all upper case letters they would modify it so that the search was in all lower case letters (the truth about cats and dogs) or in title case (The Truth About Cats and Dogs). Although ineffective, these are commendable strategies considering many computerized phrases, such as usernames and passwords, are often case sensitive. Another strategy that was repeated more then once was to drop all insignificant words from the search à la Google. Using the same title, for example, instead of searching for "truth about cats and dogs" the user would search for truth about cats dogs. An interesting strategy was used by one user when looking for the French film Les Choristes. The user originally searched for the film title in English including the initial article (The Chorus) but when that search failed to reveal the title, the user attempted to search for the title in French, but again included the initial article, in this case les. Unfortunately for the user this search also failed to find the item which indeed exists in the library's catalog.

Limitations and Further Research

Because information about specific patrons is not kept in the logs for privacy, technical, and other reasons, it was not possible to determine the experience level or patron status (undergraduate, faculty, librarians, etc.) of the user. One might assume that librarians and other library staff members have a better understanding of how to search their library's catalog, and do so more often then other users. If it were possible to eliminate their searches from the analysis, a better picture about how the average user searches the catalog could be formed.

The catalog that was studied defaulted to a title search. This may have skewed the count of the total number of intended title searches. It is possible that if another type of search was the default, the results of this study would be different. An analysis of a library catalog that uses a different default search could help clarify this.

Another limitation of this study is that it only looked at one library's search logs and the sample size was limited. Further research needs to be performed using search logs from other catalogs to confirm that the results are typical of academic libraries throughout the United States and elsewhere. A study that compares the effectiveness of the different methods utilized by various library catalogs to deal with initial articles would be useful. Another potential area of study is to investigate if different wording or instructions on the search and/or result screens would have an effect on the user strategies.

This study has found that a significant number of titles with an initial article in a library catalog were not found because the user included initial articles in the search. While
greater information literacy training might lessen this issue, it is unreasonable to assume
that any information literacy program can reach all users. Librarians and library system
designers need to design and configure more flexible systems that take this into account.
Nielsen and Pyle (1995) suggested that titles in all minor languages (any languages
beside English, French, German, and Spanish) should be indexed both with and without
initial articles. This suggestion may not go far enough. Computer storage is inexpensive
enough that indexing titles in all languages with and without initial articles is a viable
option. Further study should be performed to determine if indexing all titles with and
without initial articles would improve. Additionally, systems should be designed with
clearer error messages and hyperlinks to searches that do not contain initial articles when
a title search that contains an initial article fails. An error message for a failed search for
The truth about cats and dogs could contain a hyperlink to a title search for truth about
cats and dogs. Like Nielsen and Pyle, I hope this article will stimulate further research to
address this gap in the literature.

Notes

1 Although Hutcherson’s (2004) study did not include the term “initial articles” it did show that
even after a seven week library skills lab a significant number of undergraduate students did
recognized many common library terms including “catalog” “truncation” and “bibliography.”  Back
2 The number of searches analyzed has a confidence interval less then nine at a 95% confidence
level.  Back
3 This is computed by dividing the percentage of title searches with an initial article by the
percentage bibliographic records that contain an initial article. Back
4 Some library catalogs will place the user into the title index where the user would have found the
item. Voyager does not do this. However, if it did, the user may have been able to scroll forward in
the index and possibly been able to find the item. Back

Acknowledgements

I would like to thank Elizabeth Borland for encouraging this research, Linda Dempf for
offering some editorial changes, and the reviewers for their constructive comments and
guidance

References

catalogues bibliothèque: une étude de cas  Annual conference of the Canadian
Association for Information Science. London, Ont  Retrieved October 14, 2005,
