

Transaction Log Analysis of a Japanese Mobile Phone OPAC

Harumi MURAKAMI^{1*}, Chizu KANATA²

¹ Graduate School for Creative Cities, Osaka City University,
3-3-138, Sugimoto, Sumiyoshi, Osaka 558-8585 Japan

² Chuoystem Corporation, 1-9-15, Kitahama, Chuo, Osaka 541-0041 Japan

* Correspondence should be addressed to harumi@media.osaka-cu.ac.jp

Abstract

We analyzed the transaction log of Osaka City University's Mobile Phone Online Public Access Catalog (OPAC) and found the following results: (1) The average number of input terms was 1.4, and 92.2% of the searches only used one or two terms; (2) An automated re-search facility improved the zero-hit problem by 31.7%. We concluded that the number of input terms resembled Japanese Web-OPACs, Web searches, and mobile search engines. Two results reflect that Japanese is an agglutinative language: the number of input terms is half of western Web-OPACs and Web search engines, and an automated re-search facility is useful.

Introduction

In Japan mobile phone OPACs (Online Public Access Catalog) have been rapidly spreading in university and public libraries. Analyzing how mobile phone OPACs are used is crucial work in the so-called "Mobile 2.0" age.

Transaction log analysis has been a useful approach to investigate user search behavior. Much work has analyzed user behaviors for databases, OPACs (Peters, 1989), and Web search engines (Kashman et al., 2006) by transaction log analysis. Research into the use of mobile search engines has also been reported. However, the search behavior for mobile phone OPACs has not been thoroughly investigated.

In this paper, we analyzed the transaction log of Osaka City University's Mobile Phone OPAC (hereafter OCU-Mobile-OPAC) for 2005. The system is currently open to the public as a pilot test, and the number of users is small. However, we believe that investigating system use is valuable because scant work in the world has investigated how users utilize mobile phone OPACs.

Overview of OCU-Mobile-OPAC

OCU-Mobile-OPAC is the mobile phone OPAC of Osaka City University (Figure 1). Its features include the following: (1) a simple keyword search facility with additional narrower search by adding keywords; (2) one result page that displays a maximum of 25 results; (3) an "automated re-search facility" that breaks Japanese queries into morphons using morphological analysis and searches using these terms with an "AND" search for solving "zero-hit problems"; and (4) an Amazon search facility that displays Amazon data (customer opinions, etc.) when requested.

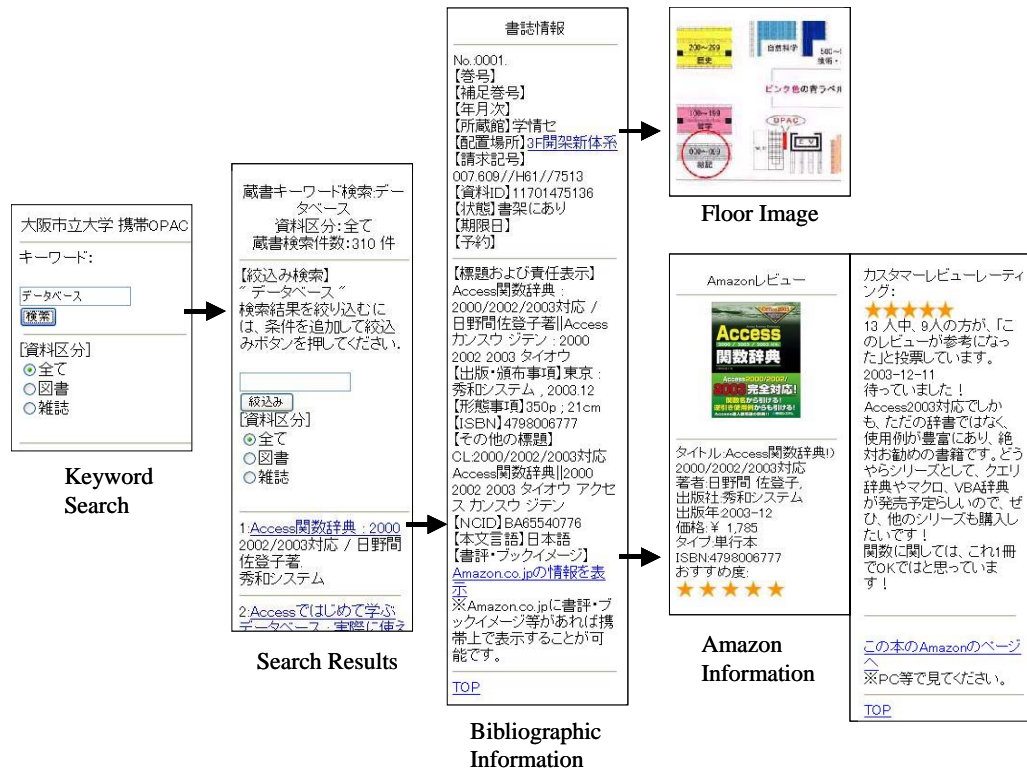
Method

Log Data

We analyzed OCU-Mobile-OPAC's transaction log for 2005. Date, time, IP addresses, mobile phone numbers, operations, etc. are recorded in the log data. The system can be used by PCs, but access logs by PCs can be removed by checking whether the data have mobile phone numbers.

Identifying Sessions

The biggest difference between logs by PCs and mobile phones is that mobile phone numbers are recorded by access from mobile phones. We exploit this advantage to identify sessions using the following procedure: If a mobile phone number of an access is identical as the previous access, and there is a 15-minute interval from the previous access, the access is included in the same session as the previous access; otherwise, the access is treated as a new session.



- (a) Keyword search screen: users can input a word or phrase and execute a search.
- (b) Search results screen: search results are displayed. Users can select a result (title), input a word or phrase into the search box and execute a narrower search, or return to the keyword search screen.
- (c) Bibliographic information screen: displays such information of the selected title. Users can display a floor image where the title is located, view Amazon information of the selected title, or return to the keyword search screen.

Figure 1: Screen example of OCU-Mobile-OPAC.

Results

Overview

The number of accesses (identical to operations) from mobile phones was 1,338. The number of identified sessions was 298. The number of different mobile phones was 43.

Figure 2 shows an overview of operations. 657 (49.1%) were search, 494 (36.9%) displayed bibliographic information, 95 (7.1%) displayed Amazon information, 50 (3.7%) were top page accesses, 27 (2.0%) displayed subsequent pages, and 15 (1.1%) displayed floor images.

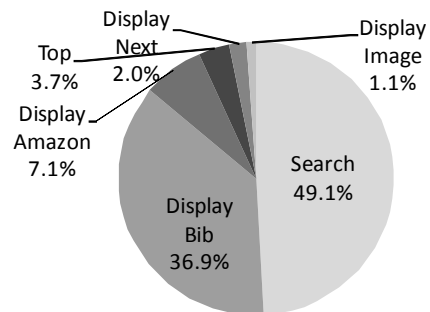


Figure 2: Overview of accesses (operations).

81.0% (132/163) of the search operations that contained more than 25 search results displayed only the first page.

Narrower searches by adding keywords were 10.5% (69/657) of the search operations.

Search Results

Figure 3 shows the ratio of number of search results. 108 (16.4%) were zero hits, 117 (17.8%) were only 1 hit, 269 (40.9%) were 2-25 hits, 140 (21.3%) were 26-999 hits, and 23 (3.5%) were more than 999 hits.

In the case of zero hits or more than 999 hits (19.9%), no result is displayed in OCU-Mobile-OPAC.

Displaying subsequent pages was performed 27 times in 26-999 hits (140 times). We therefore found that 80.7% of the operations of 26-999 hits only displayed the first page.

The automated re-search facility was performed 167 times. For 167, the number of displayed search results (i.e., 1-999 hits) is 53. In other words, the automated facility improved the zero-hit problem 31.7% (53/167).

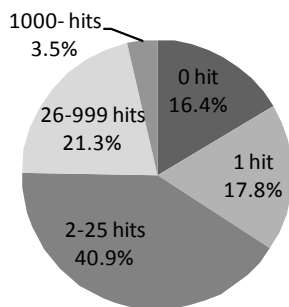


Figure 3: Number of search results.

Queries

580 were Japanese (88.8%), 57 were other languages (alphabet) (8.7%), and 16 were a mixture of Japanese and other languages (alphabet) (2.5%). English was the most common other language.

Input Terms The concept “term” differs by language. We define a term as “a character string segmented by spaces, commas, periods, or other

symbols.” The average number of input terms (terms included in queries) was 1.4.

Figure 4 shows how many terms are included in the queries. 429 (65.3%) were searched by one term. 177 (26.9%) were searched by two terms. 40 (6.1%), 5 (0.8%), and 2 (0.3%) were searched by three, four, five terms, respectively. Four queries included nothing (0.6%). In other terms, 92.2% was searched by one or two terms.

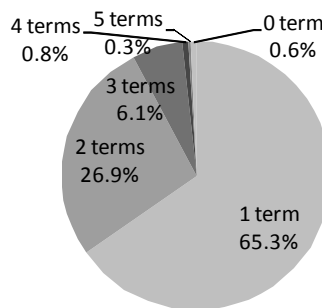


Figure 4: Input terms.

The number of terms of the longest query was 5 both in Japanese and the other languages. The number of terms of the shortest query was 1 both in Japanese and in the other languages.

Input Characters The average number of input characters (characters included in queries) was 6.5 when including spaces, and 6.1 when not including them. The number of characters of the longest query in Japanese was 22 and other languages was 32 (not including spaces). The number of characters of the shortest query was 1 both in Japanese and other languages.

Discussion

We discuss the features of the use of a Japanese mobile phone OPAC by concentrating on the input terms for comparison with other search systems. In our result, 1.4 is the average number of input terms: 1 term (65.3%) and 2 terms (26.9%).

Taneichi and Itsumura (2007) reported 1 term (66.5%), 2 terms (27.2%), 3 terms (5.9%), and more than 3 terms (0.3%) for a Japanese college library Web-OPAC in 2004-2006.

Lau & Goh (2006) reported 2 terms (34.1%), 1 term (23.6%), and 3 terms (17.3%) for a

Singaporean university library Web-OPAC. The average number of input terms was 2.86.

Inoue reports that 1.45 terms were input in Yahoo! Japan on February 2006 ("Internet Watch," (n.d.)).

According to a report by OneStat.com ("OneStat.com," (n.d.)), two or three terms are usually input in Web search engines in nine western countries. In July 2006, the frequency of input terms was 2 terms (28.9%), 3 terms (27.9%), 4 terms (17.1%), and 1 term (11.4%).

A questionnaire survey in 2006 ("Web Marketing Guide," (n.d.)) showed that 88.3% of users input one or two terms in Japanese mobile search engines: 1 term (47.2%) and 2 terms (41.5%).

Concerning input terms, our result resembles Japanese Web-OPACs, Web searches, and mobile search engines. The number of input terms is almost half of western Web-OPACs and Web search engines.

The above result reflects that Japanese is an agglutinative language. For the same reason, an automated re-search facility was also useful.

Summary

We analyzed the transaction log of Osaka City University's Mobile Phone OPAC for 2005 and found the following results: (1) The average number of input terms was 1.4, and 92.2% of the searches only used one or two terms; (2) An automated re-search facility improved the zero-hit problem by 31.7%. We concluded that the number of input terms resembled Japanese Web-OPACs, Web searches, and mobile search engines. Two results reflect that Japanese is an agglutinative language: the number of input terms is half of western Web-OPACs and Web search engines, and an automated re-search facility is useful.

References

- Web Marketing Guide. (n.d.). Retrieved May 5, 2008, from http://www.eresearch.biz/profile/pro_3/000345.html (in Japanese)
- INTERNET Watch. (n.d.). Retrieved May 5, 2008, from <http://internet.watch.impress.co.jp/cda/event/2006/04/20/11731.html> (in Japanese)
- Lau, E. P., & Goh, D. H. (2006). In search of query patterns: a case study of a university OPAC, *Information Processing and Management*, 42 (5) 1316-1329.
- Kashman, S., Spink, A., and Jansen, B. J. (2006). Web Searching on the Vivisimo Search Engine, *JASIST*, 57 (14), 1875-1887.
- OneStat.com. (n.d.). Retrieved May 12, 2008, from http://www.onestat.com/html/aboutus_pressbox4_5-search-phrases.html
- Peters, A. T. (1989). When smart people fail: an analysis of the transaction log on an online public access catalogs, *The Journal of Academic Librarianship*, 15 (5), 267-273.
- Taneichi, J. & Itsumura, H. (2007). Information seeking behavior of OPAC with contents in college libraries (in Japanese), *Proceedings of Spring Meeting of Japan Library and Information Sciences Association*, 19-22.