What a Difference a Default Setting Makes

Te Taka Keegan, Sally Jo Cunningham

University of Waikato, Computer Science Department, Hamilton, New Zealand {tetaka, sallyjo} @cs.waikato.ac.nz

Abstract. This paper examines the effect of the default interface language on the usage of a bilingual digital library. In 2005 the default interface language of a bilingual digital library was alternated on a monthly basis between Māori and English. A comprehensive transaction log analysis over this period reveals that not only did usage in a particular language increase when the default interface language was set to that language but that the way the interface was used, in both languages, was quite different depending on the default language.

Keywords: Log Analysis, Multi-Language Access

1 Introduction

Recent research on multi-language digital libraries has focused on cross-language information retrieval (CLIR)—retrieving documents written in one language through a query in a different language [1]. In this paper we consider a specific bi-language DL—the Niupepa¹ collection—and examine how the default language setting of the DL interface affects usage. We base our conclusions on the analysis of a year's site traffic and identify patterns in usage.

2 The Niupepa Digital Library Collection

The Niupepa DL (www.nzdl.org/niupepa) makes available a collection of historic Māori newspapers published between 1842 and 1933 [2]. It is a major source of historic texts of the indigenous Māori people of Aotearoa/New Zealand—just over 35,000 newspaper pages, covering 40 titles. The Niupepa DL consists of 19,106 newspaper pages written in Māori, 15,696 newspaper pages written in parallel Māori and English, 680 newspaper pages written in English, 313 English Abstracts summarizing 10 of the periodicals written in Māori, and 22 bibliographic commentaries. The newspaper pages themselves are made available as a preview image, a full size image and as a text document that has been extracted by OCR.

¹ Niupepa–Māori for newspaper

The collection is served using the Greenstone digital library software (www.greenstone.org). Three facilities are provided to access the Niupepa documents: keyword searching, browse by newspaper title and browse by issue publication date. The language of the interface may be switched between English and Māori by clicking on an icon on the home page or by accessing the Preferences page.

3 Data Collection and Definition

The default language is defined as the language that the interface to the Niupepa DL web site is displayed in when the home page (www.nzdl.org/niupepa) is requested. The default language of the interface was alternated between Māori and English in 2005. In January, March, May, July, September and November the default language was set to Māori (defMi); in February, June, August, October and December the default was set to English (defEn). There is some uncertainty about the default language setting in April due to Greenstone software upgrades that occurred at this time, and so the log data from April was discarded.

The 2005 Niupepa web logs were filtered to remove traffic from web robots, requests not recorded correctly in the log and requests from local administrators of the Niupepa DL. The resultant log was then separated into two datasets depending on the default language setting of the interface: Dataset 1 containing 93,867 defMi requests, and Dataset 2 containing 74,186 defEn requests.

Sessions were defined using an identifying argument stored as a cookie on users' computers. Although this method could not capture data from users who disabled cookies, it is more accurate in identifying individual users than simply depending on an IP address. A simple heuristic was used to delimit sessions: a session is a series of requests containing the same identifier, with no more than a 60 minute lapse between consecutive requests. The 60 minute interval is longer than a more commonly applied session interval of 30 minutes [3] primarily because of the large amount of text that are available on a Niupepa page. All the filtered requests were defined into one of three session types: single request sessions, exploratory sessions and usage sessions.

Single request sessions comprise one request submitted by a user, with no other request from that user within a 60 minute interval.

An *exploratory session* was defined as a sequence of requests submitted by a single user with no more than a 60 minute time period between each request, but no queries were submitted and no collection documents were accessed.

A *usage session* includes a sequence of requests from a single user, with no more than a 60 minute interval between successive requests, and where the sequence of requests includes one or more search, browse, or document display requests.

The usage sessions were also further divided into the usage sessions where the language of the interface was mostly set to Māori (usgMi), mostly set to English (usgEn), and a third 'bilingual' category of sessions (identified by earlier research [4]) that included a significant proportion of requests (> 20%) in both languages (usgBi). A full description of how these usage session language types were defined is given by [5] (pp 37-39).

4 Results From Web Log Data Analysis

The number of usage sessions in a particular language increased if the default language of an interface was set to that language. There were 5.8% more usgMi sessions with a defMi setting and 33.8% more usgEn sessions with a defEn setting. The usage sessions were also longer in duration and consisted of more requests if the default language setting matched the user's preferred interface language setting. The number of usage sessions defined as usgBi sessions increased by 50.6% when the default language of the interface was set to defMi.

The default language setting affected the number of English language users but not the number of Māori language users. There were 4.2 unique Māori users per day with a defEn setting and 4.5 unique Māori users per day with a defEn setting. In contrast, there were 15.8 unique English users per day with a defEn setting but only 10.1 unique English users per day with a defMi setting. This suggests that the Māori users were more comfortable with a bilingual environment and more prepared to use the interface in the non-default language.

The default language setting appeared to affect which document formats were requested. There were 11.9% more extracted text documents requested by usgMi users with a defMi setting and 13.0% more extracted text documents requested by usgEn users with a defEn setting. Subsequently, when the default language did not match the users preferred interface language the number of requests for the image forms of the documents increased by a similar percentage.

The default language setting also appeared to affect which document language types were requested by the users. With a defMi setting usgMi users had a higher preference to request Māori language newspaper pages (71.3% compared to 50.7%) and a lower preferences to request parallel Māori-English pages (34.5% compared to 14.5%). With a defEn setting UsgEn users requested English Abstracts in 17.5% of the requests but did not request them at all with a defMi setting.

For usgMi users the number of sessions with queries was highest (60.5% compared to 49.5%) with a defMi setting. For usgEn users the opposite was true; the number of sessions with queries was *lowest* (58.9% compared to 73.0%) with a defEn setting. UsgMi users used more query terms per query with a defMi setting while usgEn users used *fewer* query terms per query with a defEn setting. UsgMi users tended to use more query terms in Māori with a defMi setting than they did with a defEn setting (94.4% compared to 86.1%), whereas usgEn users appeared to use a similar ratio of query terms in Māori and English for both the defMi and defEn settings of the default interface language.

5 Summary

The default interface language setting of a bilingual/multilingual interface is very important. Not only does it affect the number of users who used the Niupepa DL, it also affected *how* these users used the DL. Usage sessions increased in number and length (both in duration and number of requests) when the default interface language matched the users' preferred interface language. Users appeared more willing to

explore and use different aspects of the interface; they accessed a different ratio of document formats, exhibited a different ratio of which language the documents were written in, and used a different ratio of Māori/English keywords. Information retrieval characteristics peculiar to a language appeared more pronounced (especially with indigenous language users) when the default language of the interface was set to that language. It was almost as if the default language of the interface was subconsciously suggesting to the user how they should undertake their information retrieval activities.

6 Conclusions

Given that the default language of the interface appears to have such a significant effect on the usage of a multilingual interface, an important question arises: which language should be set as the default interface language? Do we choose what is best for the largest number of potential users (the English interface), or do we privilege the smaller number of Māori –fluent potential users so as to support a larger social aim (the Māori language renaissance)?

The choice of a default language for a multi-language digital library collection is a more difficult decision than one perhaps originally envisions. Even a fully-fledged cross-language information retrieval system, supporting automated translation of documents and queries to the user's preferred language, can finesse some but not all of the issues raised by analysis of Digital Library collection usage. An informed decision on the appropriate default interface language clearly requires a deep understanding of the potential users for a multi-language digital library and the greater environment (political, financial, cultural, social, and so forth) in which it exists.

References

- Petrelli, D., Beauliu, M., Sanderson, M.: Observing Users, Designing Clarity: A case Study on the User-centered Design of a Cross-language Information Retrieval System. In: Journal of the American Society for Information Science and Technology 55(10), pp. 923-934. (2004)
- Apperley M. D., Keegan T. T., Cunningham S. J., Witten, I. H.: Delivering the Māori newspapers on the Internet. In: Rere Atu Taku Manu! Discovering History Language and Politics In: The Māori Language Newspapers. Auckland University Press. pp. 211-236. (2002)
- Jones, S., Cunningham, S. J., McNab, R., Boddie, S. J.: A Transaction Log Analysis of a Digital Library. In: International Journal of Digital Libraries. 3(20), pp. 152-169. (2000)
- 4. Keegan, T. T., Cunningham, S.J.: Language Preference in a Bi-language Digital Library. In: Procs of the 5th ACM/IEEE Joint Conference on Digital Libraries. pp. 174-175. (2005)
- Keegan, T. T.: Indigenous Language Usage in a Digital Library: He Hautoa Kia Ora Tonu Ai. Unpublished PhD thesis. (2007). University of Waikato, Aotearoa/New Zealand. At: adt.waikato.ac.nz/uploads/approved/adt-uow20071031.094950/public/01_Front.pdf