

Chapter 1

Design and Development Considerations for a Multilingual Digital Library

Anne R. Diekema
Utah State University, USA

ABSTRACT

Multilingual digital libraries provide access to resources in more than one language by allowing queries in one language to retrieve documents in multiple languages. By providing cross-lingual information access, multilingual digital libraries have the potential to improve international understanding, and they are an essential component of our global information society. Building a multilingual digital library presents a unique set of challenges in the areas of translation, language processing, user interface, system architecture, and project management. Perhaps the most challenging feature of the multilingual digital library is the crossing of the language barrier, which is achieved through translation, often introducing errors in the process. This chapter presents a general introduction to the challenges facing developers of multilingual digital libraries and provides references for further reading.

INTRODUCTION

This chapter presents a general introduction to the challenges facing developers of multilingual digital libraries. Each challenge is briefly described and possible approaches to these challenges from a system development point of view are provided with references to the literature. By highlighting relevant issues and pointing to relevant research in these areas this chapter provides a starting point for those interested in developing a multilingual digital library. In addition to this chapter, the reader

should also consider a review on multilinguality in the digital library (Diekema, 2012) and an article on translation approaches for the multilingual digital library (Yang & Li, 2005).

BACKGROUND

Given the existence of approximately 4,500 living languages, of which at least 30 are spoken by 30 million people or more (Edwards, 1994), multilingual digital libraries are an essential component

DOI: 10.4018/978-1-4666-2500-6.ch001

of our global information society. Unfortunately, the majority of digital libraries are monolingual (Chen & Ruiz, 2009) and contain resources in a single language, effectively restricting access to speakers of that language. Multilingual digital libraries, on the other hand, contain resources in more than one language and allow multilingual access to these materials. For example, the search interface of the World Digital Library (WDL – <http://wdl.org>) is available in seven languages: Arabic, Chinese, English, French, Portuguese, Russian, and Spanish, and contains primary source cultural content in a wide number of formats from a growing number of national libraries and other institutions (Van Oudenaren, 2010).

In addition to the WDL other large-scale international multilingual digital libraries are Europeana, the Digital Library of the Caribbean, and the International Children's Digital Library. Europeana (<http://europeana.eu>), funded by the European Union, is the European digital library and contains cultural and scientific materials from member nations (Purday, 2009); the Digital Library of the Caribbean (<http://dloc.com>) holds cultural, historical and research materials from the Caribbean (Wooldridge, et al., 2009). The International Children's Digital Library (<http://childrenslibrary.org>) includes a collection of children's literature in over 50 languages (Hutchinson, et al., 2005). Multilingual digital libraries potentially remove the language barrier for users and as a result may improve international understanding by increasing the information flow across borders and between cultures (Fox & Marchionini, 1998; Cousins, 2006; Diekema, 2012).

While constructing a monolingual digital library is complicated in and of itself (Witten, 2002), building a multilingual digital library presents a unique set of additional challenges. Perhaps the most challenging feature of the multilingual digital library is the crossing of the language barrier, which is achieved by some form of translation. Another language-related challenge for the multilingual digital library is processing

and handling different language representations or character encodings. The user interface of a multilingual digital library also requires special language considerations. Cross-Language Information Retrieval (CLIR) is the discipline that studies technical aspects involved with multilingual information access and detailed reviews of this field are provided by Oard and Diekema (1998), Kishida (2005). Additional challenges are found in system architecture development, especially when combining existing monolingual digital libraries into larger federated multilingual libraries. Besides technical challenges, multilingual digital library developers also face project management issues as extensive collaboration is almost always required in multilingual cross-cultural projects. All these challenges have consequences for the design and development of the multilingual digital library and will be discussed below.

TRANSLATION

To carry out a search in a multilingual digital library, the search query needs to be matched to documents in various languages by means of translation. In other words, for matching to take place between queries and documents they need to be in the same language. For example, in a hypothetical digital library with English and Spanish documents, a Spanish query is translated into English to retrieve English documents while the original query is used to retrieve the Spanish language documents. It is worth noting that relevant English documents might have to be translated into Spanish for Spanish-only speakers and vice versa.

Translation Approaches

There are several approaches to translation (Yang & Li, 2005). Multilingual digital library developers can choose from different translation options depending on the context of their libraries. The

10 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/design-development-considerations-multilingual-digital/72442

Related Content

A Presentation-Preserved Compositional Approach for Integrating Heterogeneous Systems: Using E-Learning as an Example

Fang-Chuan Ou Yang (2011). *International Journal of Digital Library Systems* (pp. 67-85).
www.irma-international.org/article/presentation-preserved-compositional-approach-integrating/59889

Usability Evaluation of Digital Library

Judy Jeng (2009). *Handbook of Research on Digital Libraries: Design, Development, and Impact* (pp. 278-286).
www.irma-international.org/chapter/usability-evaluation-digital-library/19891

Fuzzy-Based Answer Ranking in Question Answering Communities

B.A. Ojokoh and P.I. Ayokunle (2012). *International Journal of Digital Library Systems* (pp. 47-63).
www.irma-international.org/article/fuzzy-based-answer-ranking-in-question-answering-communities/83502

Comparing Repository Types: Challenges and Barriers for Subject-Based Repositories, Research Repositories, National Repository Systems and Institutional Repositories in Serving Scholarly Communication

Chris Armbruster and Laurent Romary (2010). *International Journal of Digital Library Systems* (pp. 61-73).
www.irma-international.org/article/comparing-repository-types/48203

First Person Singular: A Digital Library Collection that Helps Second Language Learners Express Themselves

Shaoqun Wu and Ian H. Witten (2010). *International Journal of Digital Library Systems* (pp. 24-43).
www.irma-international.org/article/first-person-singular/39035