

Chapter VIII

Design and Development of a Taxonomy Generator: A Case Example for Greenstone

Yin-Leng Theng

Nanyang Technological University, Singapore

Nyein Chan Lwin Lwin

Nanyang Technological University, Singapore

Jin-Cheon Na

Nanyang Technological University, Singapore

Schubert Foo

Nanyang Technological University, Singapore

Dion Hoe-Lian Goh

Nanyang Technological University, Singapore

ABSTRACT

This chapter addresses the issues of resource discovery in digital libraries (DLs) and the importance of knowledge organization tools in building DLs. Using the Greenstone digital library (GSDL) software as a case example, we describe a taxonomy generation tool (TGT) prototype, a hierarchical classification of contents module, designed and built to categorize contents within DLs. TGT was developed as a desktop application using Microsoft .NET Framework 2.0 in Visual C# language and object-oriented programming. In TGT, Z39.19 was implemented providing standard guidelines to construct, format, and manage monolingual controlled vocabularies, usage of broader terms, narrower terms and related terms as well as their semantic relationships, and the simple knowledge organization system (SKOS) for vocabulary specification. The XML schema definition was designed to validate against rules developed for the XML

taxonomy template, hence, resulting in the generated taxonomy template supporting controlled vocabulary terms as well as allowing users to select the labels for the taxonomy structure. A pilot user study was then conducted to evaluate the usability and usefulness of TGT and the taxonomy template. In this study, we observed four subjects using TGT, followed by a focus group for comments. Initial feedback was positive, indicating the importance of having a taxonomy structure in GSDL. Recommendations for future work include content classification and metadata technologies in TGT.

INTRODUCTION

One of the important contributions of the Web technology to modern information technology development has been the evolvement of digital libraries (DLs) that provide virtual accessibility of information resources to users at anytime, from any place where Internet access is available. DLs can be considered as the continuation of the traditional library work in digital form with the support of Internet access. DLs basically involve both primary data and manually created metadata sets. The effort in constructing a DL from scratch is huge since it actually requires an attractive and user-friendly interface with effective content management and powerful search and browsing capabilities.

A solution to this problem was implemented by creating software applications that would help to automate the DL building processes with enhanced facilities to multilingual information retrieval systems, support for interoperability protocols, and effective metadata management for diverse media formats (Witten, McNab, Boddie, & Bainbridge, 2000). During the last decade, several DL building tools have emerged aiming to provide management and distribution processes of DL collections. Some DL software packages may be subject-oriented, institution-oriented, mission-oriented, or used for digital object management. Currently, DSpace, Fedora, EPrint, and Greenstone are the most popular open-source DL software packages available that aim to empower users with capabilities to

design, build, and manage digital collections. For example, DSpace offers a platform for digital preservation for an institutional repository system while Fedora provides a service for managing digital objects. EPrint allows open access to digital contents, primarily for institutional repositories and scientific journals. Greenstone is particularly aimed at providing users with easily automated DL building processes.

Despite success in open-source software making DLs more accessible to many users, content management and metadata tagging remain important research challenges in DLs, and, in general, Web portals and Internet resources in facilitating efficient search and discovery of relevant information resources (Yan, 2004). In this chapter, we describe the design and development of a taxonomy generator supporting tagging and classification of digital resources in DLs, focusing on enhancing open-source DL software to provide more efficient search and discovery of information sources.

As a case illustration, we have selected Greenstone (<http://www.zdl.org>) to implement a taxonomy structure for better resource discovery of the digital collections. In subsequent sections, we give a brief overview of Greenstone, explain the design rationale of the taxonomy generator tool (TGT), and discuss the implementation and initial feedback of TGT. Finally, the chapter concludes with design challenges faced and lessons learned.

13 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-taxonomy-generator/19871

Related Content

Handhelds for Digital Libraries

Spyros Veronikis, Giannis Tsakonas and Christos Papatheodorou (2009). *Handbook of Research on Digital Libraries: Design, Development, and Impact* (pp. 298-307).

www.irma-international.org/chapter/handhelds-digital-libraries/19893

Design and Development of Institutional Repository Using Dspace: A Case Study of Kuvempu University, India

B. S. Biradar and Koteppa Banateppanavar (2012). *International Journal of Digital Library Systems* (pp. 1-13).

www.irma-international.org/article/design-and-development-of-institutional-repository-using-dspace/83498

Significance of Web 0.0-4.0, Semantic Web, Cloud Computing and Drupal in Concomitant to Library/ Information/ Knowledge Centres Management

Arun Kumar Chakraborty, Sumita Dey and Gopa Dasgupta (2014). *International Journal of Digital Library Systems* (pp. 1-15).

www.irma-international.org/article/significance-of-web-00-40-semantic-web-cloud-computing-and-drupal-in-concomitant-to-library-information-knowledge-centres-management/141371

Automatic Text Summarization in Digital Libraries

Shiyan Ou, Christopher S.G. Khoo and Dion Hoe-Lian Goh (2009). *Handbook of Research on Digital Libraries: Design, Development, and Impact* (pp. 159-172).

www.irma-international.org/chapter/automatic-text-summarization-digital-libraries/19879

Significance of Web 0.0-4.0, Semantic Web, Cloud Computing and Drupal in Concomitant to Library/ Information/ Knowledge Centres Management

Arun Kumar Chakraborty, Sumita Dey and Gopa Dasgupta (2014). *International Journal of Digital Library Systems* (pp. 1-15).

www.irma-international.org/article/significance-of-web-00-40-semantic-web-cloud-computing-and-drupal-in-concomitant-to-library-information-knowledge-centres-management/141371