

Knowledge Organization Systems in Canadian Digital Library Collections

Abstract: The paper reports on a study of the ways in which Canadian digital library collections make use of knowledge organization systems to support users' information search behaviour. The study identified 33 digital collections which have employed some type of knowledge organization system in their search interfaces.

Résumé : Cet article présente les résultats d'une étude sur la manière dont les systèmes d'organisation des connaissances sont utilisés par les collections des bibliothèques numériques canadiennes, afin d'assister le comportement de recherche informationnelle des utilisateurs. Cette étude a identifiée 33 collections numériques qui ont employé certains types de systèmes d'organisation des connaissances dans leurs interfaces de recherche.

Introduction

The term 'Knowledge Organization Systems' (KOS) refers to a broad range of schemes for organizing information and facilitating resource description and discovery. Knowledge organization systems include general classification schemes, special classification systems, thesauri, subject headings, taxonomies, ontologies, gazetteers and other types of standardized controlled vocabularies. Recent technological developments associated with the Internet and the World Wide Web have revitalized interest in the application of KOS in promoting knowledge management and information retrieval on the web. A review of the websites of the Government of Canada, Library and Archives Canada, Canadian Initiative on Digital Libraries, and Canada's Digital Collections indicates that academic and governmental organizations have made use of different types of KOS systems. However, there is little research on what KOS systems have been used by Canadian academic and governmental institutions on the web and how they have provided search, browse and navigation facilities based on KOS systems.

The overarching objective of the research was to evaluate the types of KOS systems and tools that have been utilized in governmental and academic digital collections in Canada.

More specific objectives of the project were:

- To identify academic and governmental digital library collections on the web that have made use of KOSs
- To determine the types of KOS systems in use
- To examine the extent to which academic and governmental organizations have provided search, browse and navigation facilities based on KOSs

Prior Research

There are a number of studies addressing issues relating to the use and applicability of knowledge organization systems such as classification schemes and thesauri in the web environment (Shiri, 2000). The Desire project (DESIRE II) investigated the use of well-

known classification schemes such as Dewey Decimal Classification (DDC), Universal Decimal Classification (UDC) and the Library of Congress Classification (LCC) to improve subject access to web resources. There are a number of subject gateways in the UK which have extensively used knowledge organization systems to support browsing and searching. Examples are Social Science Information Gateway (SOSIG), VetGate in the area of veterinary medicine, and OMNI (Organizing Medical Networked Information). A study of the application of knowledge organization systems in UK digital library collections and services revealed that the Library of Congress Subject Headings (LCSH), Dewey Decimal Classification (DDC) system and the UNESCO thesaurus were the most widely used knowledge organization systems (Nicholson et al., 2001).

Hodges and Lunau (1999) describe the digital library initiatives that the National Library of Canada (NLC) has been involved in. In particular, they address issues relating to digitization, collecting Canadian networked electronic publications, development of virtual catalogues, metadata creation, and preservation of digital resources. Issues relating to collaboration, access to digital collections and databases, and preservation are addressed in the Canadian context based on NLC's experience to date. While they point out some searching issues such as the lack of common indexes for searching and keyword searching in disparate databases and catalogues, they do not specifically address the use of KOS systems as search and browsing tools. Canadian Initiative on Digital Libraries (CIDL) is an alliance of the National Library of Canada and over 50 other Canadian libraries involved in the development of digital information resources. Member institutions include academic, public, and special libraries of all sizes. CIDL aims to promote a coordinated approach to digitization and to ensure that long-term access and preservation requirements are addressed in Canada. The CIDL provides a comprehensive list of digital library collections called Inventory of Canadian Digital Initiatives. The inventory provides search and browse facilities based on the following fields: initiative name, organization, subject, province/territory, genre, type of material, organization type, and initiative status.

Hudon and Hjartarson (2002) review a number of taxonomies and thesauri that have been used by governments of the United States, Australia, New Zealand, United Kingdom and Canada. They have noted that the Government of Canada Core Subject Thesaurus has been used by at least six government departments. Renaud (2004) describes the Government of Canada's standards and recent activities to create and manage metadata and controlled vocabularies. He refers to a number of KOS used by the Government such as Core Subject Thesaurus, Government Strategic Reference Model (GSRM), and Business Activity Structure Classification System (BASCS). He also stresses the importance of interoperability and convergence of metadata models and formats in order to achieve improved access to the electronic information. In a survey of Information and referral indexing systems in Canada, two knowledge organization systems were found to be dominant among the centres, namely the InformOntario Thesaurus and the Alliance of Information and Referral Systems (AIRS/INFO LINE) Taxonomy (InformCanada Task force on Classification).

Canadian governmental and academic institutions have also made use of and incorporated knowledge organization systems into their digital library collections. A review of the websites of the Government of Canada, Library and Archives Canada, Canadian Initiative on Digital Libraries, and Canada's Digital Collections indicates that

academic and governmental organizations have made use of different types of knowledge organization systems. However, no research has been carried out to demonstrate the types of KOS systems used and the extent to which KOS systems and tools have been used to tackle the issue of information access and retrieval. This research documents the types of KOS systems in use by Canadian governmental and academic digital library collections and it sheds light on the approaches adopted for improved information retrieval and resource discovery, in particular search, browse and navigation facilities offered within the interfaces to the knowledge organization systems.

Methodology

In order to identify as many digital library collections with KOS systems as possible the following data gathering techniques were employed:

- Review the literature related to the application of KOSs on the web, with particular reference to the Canadian context
- Identify, analyze and deep scan Canadian governmental and academic institutions websites that have made use of KOSs
- Identify and contact researchers in the area of knowledge organization to supplement the information provided by visiting the above mentioned websites
- Identify and contact people in the governmental organizations who are involved in knowledge organization and information management to gather information about their usage of KOSs
- Identify widely used Knowledge organization systems and tools
- Evaluate the interfaces enhanced with KOSs in terms of search, browse and navigation facilities. The task involved an investigation of whether and how KOSs have been incorporated into the interfaces of the governmental and academic websites

There were a number of important web-based resources that provided useful information on the types of controlled vocabularies used in Canada. The following websites were among the main resources:

- Government of Canada
- Library and Archives Canada --Thesauri and Controlled Vocabularies
- Canadian Initiative on Digital Libraries and
- Canada's Digital Collections

Results

In total 33 digital collections were identified to have made use of controlled vocabularies. 20 of the collections were academic and the remaining 13 were governmental. Appendix A provides the list of digital library collections surveyed in this study.

Table 1 shows the type of controlled vocabularies and the number of collections that have used a particular controlled vocabulary. It should be noted that some collections used more than one subject scheme, therefore the total number of schemes exceeds 33.

Controlled vocabularies used	No. of collections
Library of Congress Subject Headings (LCSH)	15
Canadian Subject Headings (CSH)	4
Dewey Decimal Classification (DDC)	3
Art & Architecture Thesaurus (AAT)	3
<u>Répertoire de Vedettes-Matière (RVM)</u>	1
Locally developed thesaurus/taxonomy	11
Source index (i.e. index of digitized book)	4

Table 1. Cumulative frequency of the Controlled vocabularies used in digital collections

As can be seen from Table 2 LCSH, DDC, Canadian Subject Headings and AAT have been used by 25 collections. 30% of the collections had locally developed thesauri and taxonomies. Four digital collections provided a back of the book type index to improve subject access.

One of the objectives of this study was to evaluate the ways in which the surveyed digital library collections provided search, browse and navigation facilities based on the controlled vocabularies used. A number of interface features were evaluated in terms of allowing the user to perform searches or browse the collections. These features were searching, browsing, display and navigation and multiple access points. The interface features of the knowledge organization systems are outlined in Appendix B. The interfaces to the digital collections surveyed provided various searching facilities. For instance, some offer simple search boxes while others provide controlled vocabulary assisted search options. One of the major issues of the surveyed sites was the fact that the browsing capabilities of the database were either limited or hidden to the user. In fact, thirteen of the collections did not have a browse button on the front page of the site, although eleven of these thirteen do, in fact, have browsing capabilities. Also, even the sites that did have a browsing button, many of them had ambiguous names such as “Indexes”, “Author & Subject Index” and “Scan Index”.

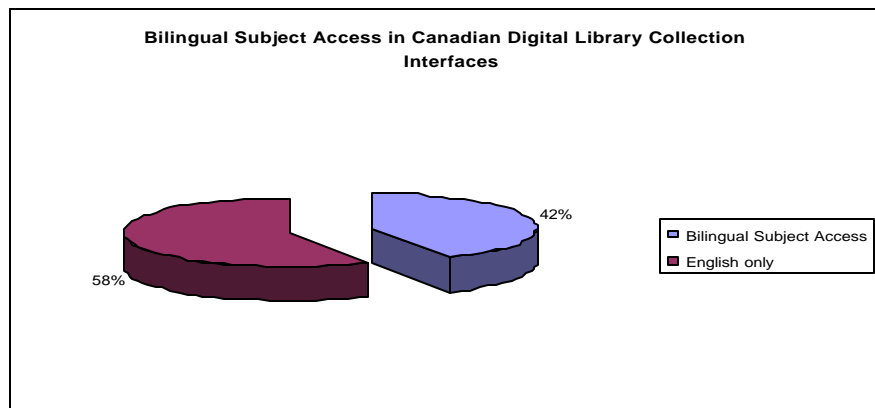


Figure 1. Bilingual Subject Access in Canadian Digital Library Collection Interfaces

Design implications and suggestions

- Subject access features should cater for search term selection, query formulation, reformulation and modification
- Explicit controlled vocabulary-based search, browse and navigation facilities should be provided
- The interface features such as navigation and browsing facilities should be intuitive, uncluttered and easy to understand
- Bilingual subject access should be supported at the interface level
- Research can be carried out to investigate cross searching and interoperability among digital library collections using widely used schemes
- Hierarchical as well as alphabetical displays of terminologies should be provided to support users' different information searching behaviours.
- A number of usability issues such as labeling and navigational styles should be addressed.

Conclusion

- Search interface design plays a key role in allowing users to explore subject access facilities in order to formulate, expand or enhance their search queries.
- Searching, browsing and navigation facilities as well as bilingual features of the survey digital collections call for improvements.
- One major challenge in the design of digital library interfaces is to make provisions for the explicit display of and easy access to controlled vocabularies to support users' query formulation process.

References

Canada's digital Collections: <http://collections.ic.gc.ca/E/view.html>

Canadian Initiative on Digital Libraries: <http://www.collectionscanada.ca/cidl/cidle.html>

Government of Canada: http://canada.gc.ca/main_e.html

DESIRE project: <http://www.lub.lu.se/desire/radar/reports/D3.2.3/> (accessed October 20, 2004).

Hodges, D., Lunau, C.D. (1999) "The National Library of Canada's digital library initiatives", *Library Hi Tech*, vol. 17, no. 2, pp. 152-164

Hudon, M., Hjartarson, F. (2002) "Governments meet people: developing metathesauri in the framework of government online initiatives", *Advancing knowledge: Expanding horizons for information science : Proceedings of the 30th Annual Conference of the Canadian Association for Information Science, 30 May - 1 June 2002*, Toronto, Canada, pp. 46-60.

Inventory of Canadian Digital Initiatives.

<http://www.collectionscanada.ca/initiatives/index-e.html>

Library and Archives Canada
<http://www.collectionscanada.ca/index-e.html>

Library and Archives Canada Thesauri and Controlled Vocabularies
<http://www.collectionscanada.ca/8/4/r4-280-e.html>

Nicholson, D. et al. (2001) HILT: High Level Thesaurus Project: Final Report. Available at:
<http://hilt.cdrl.strath.ac.uk/Reports/Documents/HILTfinalreport.doc>

OMNI: <http://omni.ac.uk/>

Renaud, G. (2004) "Metadata and controlled vocabularies in the Government of Canada: a situational analysis". *International Conference on Dublin Core and Metadata Applications 2004*, http://purl.org/metadatasearch/dcconf2004/papers/Paper_20.pdf

Shiri, A., Revie, C. (2000) "Thesauri on the web: current developments and trends". *Online Information Review*, vol. 24, no. 4, pp. 273-279.

SOSIG: <http://www.sosig.ac.uk/>

VetGate: <http://vetgate.ac.uk/>

Appendix A: Digital Library Collections

Aboriginal Canada Portal
<http://www.aboriginalcanada.gc.ca/acp/site.nsf/en-frames/index.html>

AgriWeb Canada
<http://www.agr.gc.ca/agriweb/index.cfm>

Alberta Folklore and Local History Collection
<http://folklore.library.ualberta.ca/index.cfm>

Alberta InSight
<http://asalive.archivesalberta.org:8080/access/asa/photo/>

Alberta InWord
<http://asalive.archivesalberta.org:8080/access/asa/documents/>

Alberta Online Medical History Collection
<http://www.ourfutureourpast.ca/medhist/>

Arctic Science and Technology Information System (ASTIS)
<http://www.aina.ucalgary.ca/astis/>

The Art Collection
<http://inmagic.lib.ucalgary.ca/dbtw-wpd/nam/index.html>

Bio-DiTRL
<http://bio-ditrl.sunsite.ualberta.ca/>

Canadian Architect and Builder
<http://digital.library.mcgill.ca/cab/index.htm>

Canadian Health Network
<http://www.canadian-health-network.ca/>

Canadian Illustrated News
<http://www.collectionscanada.ca/cin/>

Canadian Information by Subject
<http://www.collectionscanada.ca/caninfo/>

The Champlain Society Digital Collection (Toronto)
<http://link.library.utoronto.ca/champlain/search.cfm>

Early Canadiana Online
<http://www.canadiana.org/eco/english/index.html>

Grande Prairie in Words and Pictures: Isabel Campbell photographic collection
<http://www.ourfutureourpast.ca/grandeprairie/campbell.asp>

IDRC Development Research Information System
<http://idrinfo.idrc.ca/scripts/minisa.dll/144/IDRIS?DIRECTSEARCH>

IMAGES: The IDRC Image Library :
http://idrinfo.idrc.ca/scripts/minisa.dll/144/IDRC_IMAGES?DIRECTSEARCH

IRC Publications Database
<http://irc.nrc-cnrc.gc.ca/ircpubs/>

LeughSeo: Catalogue of Gaelic Books in Cape Breton County Libraries
<http://cbri.ca/search/leughseo.html>

McGill University David Hume Collection
<http://digital.library.mcgill.ca/hume/>

McGill University Napoleon Collection
<http://digital.library.mcgill.ca/napoleon/>

National Gallery of Canada Bulletin & Annual Bulletin, 1963-1985
<http://collections.ic.gc.ca/bulletin/>

New Brunswick Commissions of Enquiry
<http://www.lib.unb.ca/Texts/NBHistory/Commissions/index.htm>

The Northwest Resistance
<http://library.usask.ca/northwest/>

Our Roots
<http://www.ourroots.ca/>

Peel's Prairie Provinces
<http://peel.library.ualberta.ca/index.php>

Resources for Aboriginal Studies
<http://library.usask.ca/native/>

Saskatchewan News Index
<http://library.usask.ca/sni/>

The SFU Library Editorial Cartoons Collection
<http://edocs.lib.sfu.ca/projects/Cartoons/index.htm>

Ukrainian Folklore Archives Photograph Collection

http://129.128.116.48:8890/photo_archives/index.html

Urban Plans Collection

<http://digital.library.mcgill.ca/urbanplanning/>

Waldren Studios: Images of Nova Scotia, 1870-1940

<http://www.library.dal.ca/archives/waldren/index.htm>

Appendix B: Interface features of KOSs

Title of collection/service	Search button on front page	Search box on front page	Browse button on front page	Clickable subject headings in record	Hierarchical display of subject headings	Alphabetical list of subject headings	Display number of hits for each subject heading
Aboriginal Canada Portal	Yes	Yes	No	No	Yes	No	No
AgriWeb Canada	Yes	Yes	Yes	No	Yes	Yes	No
Alberta Folklore and Local History Collection	Yes	Yes	Yes	Yes	No	Yes	No
Alberta InSight	Yes	Yes	Yes	Yes	No	Yes	Yes
Alberta InWord	Yes	Yes	Yes	Yes	No	Yes	Yes
Alberta Online Medical History Collection	Yes	No	Yes	Yes	No	Yes	No
Arctic Science and Technology Information System	Yes	Yes	No	Yes	No	Yes	No
Art Collection, The	Yes	No	No	Yes	No	Yes	Yes
Bio-DiTRL	Yes	Yes	Yes	Yes	Yes	No	Yes
Canadian Architect and Builder	Yes	No	Yes	No	No	Yes	No
Canadian Health Network	Yes	Yes	Yes	No	No	Yes	No
Canadian Illustrated News	Yes	Yes	Yes	No	No	Yes	No
Canadian Information by Subject	Yes	No	Yes	No	Yes	Yes	Yes
Champlain Society Digital Collection (Toronto)	Yes	Yes	Yes	Yes	No	Yes	No
Early Canadian Online	Yes	No	No	No	No	Yes	Yes
Grande Prairie in Words and Pictures: Isabel Campbell photographic collection	Yes	Yes	Yes	No	No	Yes	No
IDRC Development Research Information System	Yes	Yes	No	Yes	No	No	No

Title of collection/service	Search button on front page	Search box on front page	Browse button on front page	Clickable subject headings in record	Hierarchical display of subject headings	Alphabetical list of subject headings	Display number of hits for each subject heading
IMAGES: The IDRC Image Library	Yes	Yes	No	Yes	No	Yes	No
IRC (Institute for Research in Construction) Publications Database	Yes	Yes	Yes	No	Yes	No	No
LeughSeo: Catalogue of Gaelic Books in Cape Breton County Libraries	Yes	No	Yes	No	No	Yes	No
McGill University David Hume Collection	Yes	No	Yes	No	No	Yes	No
McGill University Napoleon Collection	Yes	No	Yes	Yes	No	Yes	No
National Gallery of Canada Bulletin & Annual Bulletin, 1963-1985	No	No	Yes	No	No	Yes	No
New Brunswick Commissions of Enquiry	Yes	No	Yes	No	No	No	No
Northwest Resistance, The	No	No	No	Yes	No	Yes	Yes
Our Roots	Yes	Yes	Yes	Yes	No	Yes	No
Peel's Prairie Provinces	Yes	Yes	No	Yes	No	Yes	Yes
Resources for Aboriginal Studies	Yes	No	No	Yes	No	Yes	Yes
Saskatchewan News Index	Yes	No	No	Yes	No	Yes	Yes
SFU Library Editorial Cartoons Collection	Yes	Yes	No	Yes	No	Yes	No
Ukrainian Folklore Archives Photograph Collection	No	No	No	No	Yes	No	No
Urban Plans Collection	No (in a drop-down box)	No	No	No	No	Yes	No
Waldren Studios: Images of Nova Scotia, 1870-1940	Yes	No	Yes	No	No	No	No