

JCDL 2006: Opening Information Horizons
Metadata Tools for Digital Resource Repositories Workshop
 June 15, 2006, Chapel Hill, NC, USA

Product	Exhibitor		Institution
CalEDLN	Deanne DiPietro	GIS & Information Services Program Manager	Sonoma Ecology Center
CAT	Anne R. Diekema	Assistant Research Professor	Center for Natural Language Process- ing, Syracuse University
Connexion	Jon Fausey	Software Architect	Online Computer Library Center, OCLC Inc.
CONTENTdm	Claire Cocco	Product Manager	DiMeMa Inc.
DCS	John Weatherly	DLESE Software Engineer	University Corporation for Atmospheric Research (UCAR)
Greenstone	David Bainbridge	Department of Computer Science	University of Waikato, New Zealand
iVia	Jason Scheirer	Programmer, INFOMINE Project	iVia
jOAI	John Weatherley	DLESE Software Engineer	University Corporation for Atmospheric Research (UCAR)
Millennium	Spenser Thompson	Marketing Specialist	Innovative Interfaces Inc.
Seamark Navigator	Brad Allen	Founder and CTO of Siderean Soft- ware, Inc	Siderean Software Inc
TM4L	Darina Dicheva Christo Dichev	Professor of Computer Science Associate Professor of Computer Sci- ence	Winston-Salem State University

JCDL 2006: Opening Information Horizons
Metadata Tools for Digital Resource Repositories Workshop
June 15, 2006, Chapel Hill, NC, USA

CalEDLN: California Environmental Digital Library

Deanne DiPietro, GIS & Information Services Program Manager
(707) 996-0712 Ext. 114
deanne@sonomaecologycenter.org

Sonoma Ecology Center

Product Description

The California Environmental Digital Library Network (CalEDLN) is a project dedicated to the development of an open network of digital libraries that share data and information about California's natural environment. Partners are the California Resources Agency's CERES Program, the UC Davis Information Center for the Environment, an Sonoma Ecology Center. A pilot is being implemented for the resource management community in the North San Francisco Bay area where collaborators are working on making existing meta-databases interoperable and semantically linked as part of their Northbay Environmental Information Commons project.

Sharing and linking metadata about projects and information resources in this distributed environment presents us with many challenges. Among them are those of providing access to thematic and geographic controlled vocabularies as well as names of people and organizations in order to convey the semantics of the source data and support the inter-linkages needed for the network.

A standard for data exchange based on the FGDC and Dublin Core metadata standards has been developed using URIs to reference controlled vocabularies posted as SKOS encoded vocabularies. A web service API (application programming interface) allows for interaction between thesauri and metadata management or search interface software using JSON to transfer data objects. The API supports term matching and thesaurus functionality for thematic keywords, California place names, resource types, and person/organization names. A demonstration web-based metadata editing application that utilizes the API to access and display several community-wide thesauri will be shown. It is anticipated that multiple, domain-specific thesauri will be increasingly needed to grow the network, and we hope to provide for this need using these techniques.

The CalEDLN and Northbay Commons projects are open-source collaborations, and these tools are being developed with the intention of supporting innovation in end-user technology as well as contribution and community management of distributed vocabularies.

Related Resources

<http://caledln.org>

<http://northbaycommons.org>

CAT: The Metadata Tool for Standards Assignment

Anne R. Diekema
diekemar@syr.edu
<http://casaa.cnlp.org/casaa>

Center for Natural Language Processing, Syracuse University

Product Description

CAT is a Tool that assists collection providers, catalogers and teachers in assigning content standards by providing suggestions of relevant standards. The Tool makes suggestions which are reviewed by the individual adding the resource to the NSDL. The user then selects and approves the final assignment. As well, if the individual adding the resource would like to consider other standards, the system allows the user to browse a hierarchical list of available standards and select and add their preferred standards to the list of suggested standards. The system stores these human-vetted assignments and utilizes machine learning techniques in order to improve its own auto-suggestions. This learning can take place at the single cataloger level, or at the organizational level where all assignments from multiple catalogers in an organization are aggregated. Special CAT features are:

NLP

CAT uses Natural Language Processing (NLP) on the content of the educational resources and standards. Through NLP, the system is able to process both the standard and the educational resource with part-of-speech tags, and then bracket verb and noun phrases, categorize terms and phrases, and find synonyms for them in order to facilitate more accurate matching between the resource and the standards.

User Relevance Feedback “More Like This”

Once the system has provided suggestions, the user can select one or more standards from the results and, if they wish, ask the system to find more standards similar to those selected.

Facilitated Standards Navigation “Browse Standards”

Typically, there are a vast number of standards in a single standards document such as the National Council of Teachers of Mathematics and National Science Education Standards. Navigation within this document is cumbersome at best. But CAT provides an easily navigable standards tree for users to browse the standards and ease their cognitive effort.

Machine Learning

We further utilize the human-vetted assignments in auto-teaching CAT how to make better assignments. Thus CAT incorporates a continuous assignment quality improvement through a custom built instance-based machine learning algorithm.

API

The capabilities of CAT are abstracted to an API which makes use of a TextTagger server and an NLP Information Extraction engine to process learning resources and to suggest relevant standards. The vetted relevance judgments are persisted in a relational database. The API is currently used by the Digital Library for Earth System Education (DLESE) to integrate CAT with DLESE's cataloging system (DCS).

JCDL 2006: Opening Information Horizons
Metadata Tools for Digital Resource Repositories Workshop
June 15, 2006, Chapel Hill, NC, USA

OCLC Connexion



Johnathan Fausey
fausey@oclc.org

Online Computer Library Center, OCLC Inc.

Jon has worked in research and development for seventeen years at OCLC Online Computer Library Center, where he is currently a Software Architect. His recent work has focused on OCLC's online cataloging system, Connexion, and associated services. More recently, spinning-off or integrating new web services that provide functionality such as bibliographic and authority record validation or automated extraction and creation of metadata from electronic resources. Jon's email address is fausey@oclc.org.

Product Description

Libraries have a growing need to represent digital resources in local catalogs, as well as to share them via Open WorldCat and WorldCat. OCLC Connexion's Extract Metadata functionality currently allows catalogers to extract metadata from PDF, HTML, and Word documents and to create (infer) additional metadata as the basis of a bibliographic workflow.

This helps automate cataloging of such resources by putting information from the digital resource directly into the workflow. This saves time and reduces errors. It also allows libraries to integrate their electronic resources into their collections and catalogs with a single cataloging interface.

Related Resources

OCLC Connexion Site
<http://www.oclc.org/connexion/>

JCDL 2006: Opening Information Horizons
Metadata Tools for Digital Resource Repositories Workshop
June 15, 2006, Chapel Hill, NC, USA



Claire Cocco, Product Manager
860.216.5507
claire@dimema.com

DiMeMa Inc.

Claire Cocco has been working on Web-based content delivery systems and applications since 1994. She joined DiMeMa, Inc. in 2002 and is the CONTENTdm Product Manager.

Product Description

CONTENTdm® Digital Collection Management Software, developed by DiMeMa, Inc., provides a complete set of tools for organizing, managing, publishing and searching digital collections over the Internet. The most powerful and flexible digital collection management package on the market today, CONTENTdm handles all media types—documents, newspapers, PDFs, images, video, and audio files. CONTENTdm is used by hundreds of libraries, universities, government agencies, museums, corporations, historical societies, and a host of other organizations to support thousands of diverse digital collections.

Related Resources

For more information about CONTENTdm see www.contentdm.com.

JCDL 2006: Opening Information Horizons
Metadata Tools for Digital Resource Repositories Workshop
June 15, 2006, Chapel Hill, NC, USA

DCS: a Versatile Tool for Metadata Creation and Management

John Weatherley DLESE Program Center, University Corporation for Atmospheric Research (UCAR)

<http://www.dlese.org> jweather@ucar.edu (e-mail)
PO Box 3000 Boulder, CO 80307-3000e
(303) 497-2680 (tel) / (303) 497-8336 (fax)

John holds an MS in Computer Science. His thesis research examined how software systems can support the peer-review of complex digital learning objects. John has significant expertise in information retrieval, web services, and digital library architectures, and has received widespread recognition for his contributions to digital library technologies, specifically his Open Archives Initiative (OAI) software, the DLESE search web service, and developers' web service toolkit and documentation. In 2002, the Joint Conference on Digital Libraries nominated one of John's papers for the Vannevar Bush Best Paper award. He is a full-time software engineer at the DPC, serving as the lead for the DLESE Discovery Services team.

Product Description

The Digital Library for Earth System Education (DLESE - www.dlese.org) has developed the DLESE Collection System (DCS) as a tool for creating, editing, and managing metadata records and collections. Metadata editing interfaces are automatically generated from XML schema, allowing the DCS to support virtually any metadata framework expressed as an XML schema. The metadata editing interfaces provide incremental validation of input as well as contextualized "best practice" information for each field. Individual metadata records are organized into collections and support is provided for searching and browsing over the collections, as well as for collection management workflows.

The DCS provides three types of web services that enable the tool to interoperate with other digital libraries and tools: 1) an OAI provider that enables metadata to be harvested; 2) a Search service that enables the construction of customized end-user search and discovery interfaces; and 3) a Repository service that enables authorized 3rd party applications to insert metadata into the DCS programmatically.

The DCS is highly configurable to enable significant tailoring of behavior without requiring programming skills. The system is currently used by DLESE as the primary collection management tool. A prototype version of the DCS aids catalogers in the assignment of educational standards to resources by automatically suggesting a list of standards that may align with the resource. This feature uses the CAT Suggestion Server (exhibited in this workshop), developed by the Center for Natural Language Processing at Syracuse University.

Related Resources

A "sandbox" version of the DCS is accessible at <http://dcs.dlese.org/preview>. For information about obtaining a DCS, please visit http://www.dlese.org/libdev/dcs_overview.html.

JCDL 2006: Opening Information Horizons
Metadata Tools for Digital Resource Repositories Workshop
June 15, 2006, Chapel Hill, NC, USA

Greenstone Digital Library System

David Bainbridge, Department of Computer Science
davidb@cs.waikato.ac.nz
www.greenstone.org

University of Waikato, New Zealand

Product Description

Greenstone is a suite of software for building and distributing digital library collections. It is not a digital library but a tool for building digital libraries. It provides a new way of organizing information and publishing it on the Internet in the form of a fully-searchable, metadata-driven digital library. It has been developed and distributed in cooperation with UNESCO and the Human Info NGO in Belgium. It is open-source, multilingual software, issued under the terms of the GNU General Public License. Its developers received the 2004 IFIP Namur award for “contributions to the awareness of social implications of information technology, and the need for an holistic approach in the use of information technology that takes account of social implications.”

Platforms. Greenstone runs on all versions of Windows, and Unix, and Mac OS-X. It is very easy to install. For the default Windows installation absolutely no configuration is necessary, and end users routinely install Greenstone on their personal laptops or workstations. Institutional users run it on their main web server, where it interoperates with standard web server software (e.g. Apache).

Interoperability. Greenstone is highly interoperable using contemporary standards, It incorporates a server that can serve any collection over the Open Archives Protocol for Metadata Harvesting (OAI-PMH), and Greenstone can harvest documents over OAI-PMH and include them in a collection. Any collection can be exported to METS (in the Greenstone METS Profile, approved by the METS Editorial Board and published at <http://www.loc.gov/standards/mets/mets-profiles.html>), and Greenstone can ingest documents in METS form. Any collection can be exported to DSpace ready for DSpace’s batch import program, and any DSpace collection can be imported into Greenstone.

Interfaces. Greenstone has two separate interactive interfaces, the Reader interface and the Librarian interface. End users access the digital library through the Reader interface, which operates within a web browser. The Librarian interface is a Java-based graphical user interface (also available as an applet) that makes it easy to gather material for a collection (downloading it from the web where necessary), enrich it by adding metadata, design the searching and browsing facilities that the collection will offer the user, and build and serve the collection.

Metadata formats. Users define metadata interactively within the Librarian interface. These metadata sets are predefined: Dublin Core (qualified and unqualified), RFC 1807, NZGLS (New Zealand Government Locator Service), AGLS (Australian Government Locator Service)

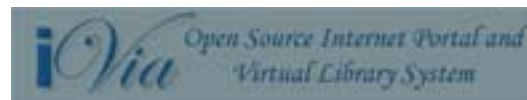
New metadata sets can be defined using Greenstone’s Metadata Set Editor. “Plug-ins” are used to ingest externally-prepared metadata in different forms, and plug-ins exist for XML, MARC, CDS/ISIS, ProCite, BibTex, Refer, OAI, DSpace, METS

Document formats. Plug-ins are also used to ingest documents. For textual documents, there are plug-ins for PDF, PostScript, Word, RTF, HTML, Plain text, Latex, ZIP archives, Excel, PPT, Email (various formats), source code

For multimedia documents, there are plug-ins for Images (any format, including GIF, JIF, JPEG, TIFF), MP3 audio, Ogg Vorbis audio, and a generic plug-in that can be configured for audio formats, MPEG, MIDI, etc.

JCDL 2006: Opening Information Horizons
Metadata Tools for Digital Resource Repositories Workshop
June 15, 2006, Chapel Hill, NC, USA

iVia: Internet Portal and Virtual Library System



Jason Scheirer Programmer, INFOMINE Project
<http://ivia.ucr.edu/>
jason.scheirer@gmail.com

iVia

Jason Scheirer graduated from the University of California, Riverside in 2005 with a degree in Linguistics and Computer Science. He researches computer simulations of linguistic phenomena and statistical natural language processing. He currently works for the iVia/Infomine project on improving classification accuracy in DataFountains.

Product Description

Data Fountains is an open source (LGPL and GPL), self-service resource discovery, metadata generation and rich text extraction utility for building collections of digital, web-based resources. Its goal is not only to discover data but to automatically produce high-quality metadata about the resources it discovers. It also features the ability to crawl in focused mode for the improved discovery of new, topical web resources and unfocused mode for the long-term indexing and generation of full collections and web sites.

This presentation outlines the techniques described in the paper *Developing Practical Automatic Metadata Assignment and Evaluation Tools for Internet Resources* (Paynter 2005). The automatic assignment of metadata to internet resources is a large problem of immediate practical use to managers of large digital libraries and metadata collections. The generation of metadata is a complicated process that includes different evaluation and generation criteria for several different classes of data to be generated.

The iVia Virtual Library Software Project approaches this by creating an evaluation workflow using incremental, hybridized human evaluation and objective statistical evaluation tools rather than a one-time evaluation of the generation tools. Repository maintainers using the automatic tools can assist in improving the machine's results and increase confidence in the results of automatic metadata generation and extraction tools.

Related Resources

Steve Mitchell, Margaret Mooney, Julie Mason, Gordon W. Paynter, Johannes Ruscheinski, Artur Kedzierski and Keith Humphreys (2003) [iVia Open Source Virtual Library System](#). *D-Lib Magazine* 9 (1).

Gordon Paynter (2005) [Developing Practical Automatic Metadata Assignment and Evaluation Tools for Internet Resources](#). Proc. Fifth ACM/IEEE Joint Conference on Digital Libraries, Denver, Colorado, June 7-11, 2005, ACM Press, pp. 291-300.

JCDL 2006: Opening Information Horizons
Metadata Tools for Digital Resource Repositories Workshop
June 15, 2006, Chapel Hill, NC, USA

jOAI: The low-barrier OAI data provider and harvester

John Weatherley DLESE Program Center, University Corporation for Atmospheric Research (UCAR)
<http://www.dlese.org> jweather@ucar.edu (e-mail)
PO Box 3000 Boulder, CO 80307-3000e
(303) 497-2680 (tel) / (303) 497-8336 (fax)

John holds an MS in Computer Science. His thesis research examined how software systems can support the peer-review of complex digital learning objects. John has significant expertise in information retrieval, web services, and digital library architectures, and has received widespread recognition for his contributions to digital library technologies, specifically his Open Archives Initiative (OAI) software, the DLESE search web service, and developers' web service toolkit and documentation. In 2002, the Joint Conference on Digital Libraries nominated one of John's papers for the Vannevar Bush Best Paper award. He is a full-time software engineer at the DPC, serving as the lead for the DLESE Discovery Services team.

Product Description

jOAI is a low-barrier Java OAI data provider and harvester tool designed to be accessible to a wide range of individuals and institutions. jOAI is a Web application that runs in Apache Tomcat or other Java servlet containers. Users set up and interact with the software through a simple Web-based user interface that does not require prior knowledge of or familiarity with the OAI protocol. The data provider portion of the software exposes selected XML files from the local machine as an OAI repository. Files in the data provider can be split into OAI sets, which can be defined using the location of the file on the machine, the XML format of the file or by the content found in the files. In addition, text-based searches can be performed over the content of the files using the Web interface or through a REST-based Web service.

The harvester portion of the tool is used to save harvested files to the local machine and allows for harvests to be performed manually or automatically at regular intervals. Each set of harvested files is also packaged to a Zip file, which can be downloaded and opened directly through the Web interface. Because the tool operates over files on the local machine, it is equally suited for use by both small and large institutional and individual repositories. jOAI is developed by the University Corporation for Atmospheric Research (UCAR), DLESE Program Center and is available through an open source license.

Related Resources

jOAI is available at: http://www.dlese.org/dds/services/joai_software.jsp

The DLESE services portal: <http://www.dlese.org/dds/services/>

The Open Archives Initiative: <http://www.openarchives.org/>

JCDL 2006: Opening Information Horizons
Metadata Tools for Digital Resource Repositories Workshop
June 15, 2006, Chapel Hill, NC, USA

Innovative Interfaces Inc. - Millennium

INNOVATIVE
i n t e r f a c e s

Katja Moos, Customer Sales Consultant
kmoos@iii.com
5850 Shellmound Way, Emeryville, CA 94608
Tel: 919/384-7147 / Fax: 928/222-0452

Innovative Interfaces Inc.

Product Description

Libraries want to provide integrated access to their digital special collections, while bringing to bear the full descriptive potential of available metadata schema. Libraries around the world are using Millennium, Innovative's integrated library technology platform, to make this a reality.

Libraries may choose to import XML metadata into the library's online catalog from an external repository or a content management system. Others choose to build digital archival metadata within Millennium, creating metadata with EAD or Qualified Dublin Core XML.

The presentation will show how Millennium's XML Harvester and MetaData Builder can be used to expand both the scope and level of integration offered by today's digital library, with examples from implementing libraries.

Related Resources

Innovative Interfaces Inc: <http://www.iii.com/>

JCDL 2006: Opening Information Horizons
Metadata Tools for Digital Resource Repositories Workshop
June 15, 2006, Chapel Hill, NC, USA

Seamark Navigator: Faceted Navigation of User-Generated Metadata



Bradley Allen
ballen@siderean.com / <http://www.siderean.com>
(310) 647 5610 / Skype: bpa777

Siderean Software Inc.

Bradley P. Allen is founder and CTO of Siderean Software, Inc., a leading provider of digital navigation solutions. Mr. Allen began his career as a member of the research staff at Carnegie-Mellon's Robotics Institute. As a senior member of the technical team at Inference Corporation he created CBR Express, one of the first case-based reasoning solutions for customer support. He was founder and CTO at Limbex Corporation, where he created WebCompass, an Internet search assistant that won Best Of Show at COMDEX Fall '95. Mr. Allen was also founder of TriVida Corporation, a Web site personalization application services provider, serving as CTO until TriVida's acquisition by Be Free, Inc. in 2000. Mr. Allen earned a BS in applied mathematics from Carnegie-Mellon University in 1982.

Product Description

Faceted navigation depends on metadata to organize and structure access to information. Traditionally, this metadata has come from relational databases, content management repositories and/or labor-intensive processes for the manual generation of subject and asset metadata by authors and/or specialists. However, applications are now appearing that instead involve end users in the generation of metadata, either through the creation of folksonomies through end user tagging or through the accumulation of usage metadata based on end user behavior. We discuss applications and work processes that blend these types of user-generated metadata with faceted navigation.

Siderean's Seamark Navigator reveals the full scope of digital assets and cataloged items available to enterprise users and customers, presenting a discovery and precision search framework for navigating to the most pertinent choices. Ideally suited for a wide variety of markets and organizations, Seamark Navigator improves overall business performance by providing a clear, contextual map of available digital materials, quickly directing users to all relevant assets or goods, leading to better individual and group decisions, content usage and purchasing.

Related Resources

Siderean Software: <http://www.siderean.com/>

JCDL 2006: Opening Information Horizons
Metadata Tools for Digital Resource Repositories Workshop
June 15, 2006, Chapel Hill, NC, USA

TM4L: Creating and Browsing Topic Map-based Learning Repositories

Darina Dicheva, Professor of Computer Science
dichevad@wssu.edu / <http://myweb.wssu.edu/dichevad/>
3206 Elva J. Jones Computer Science Building
Winston-Salem, N.C. 27110

Winston Salem State University

Christo Dichev, Associate Professor of Computer Science
dichevc@wssu.edu / <http://gorams.wssu.edu/faculty/dichevc/>
3206 Elva J. Jones Computer Science Building
Winston-Salem, N.C. 27110

Winston Salem State University

Product Description

Digital course libraries are educational Web applications that contain instructional materials to assist students' learning in a specific discipline. They play a vital role in out-of-class learning, especially in project-based and problem-based learning, as well as in lifelong learning. TM4L is an e-learning environment developed at Winston-Salem State University in North Carolina, which enables the creation, maintenance, and use of ontology-aware learning repositories based on Topic Maps. The system provides support in designing and maintaining a conceptual structure of a collection of resources through its functionality for editing, browsing, and combining such structures, coupled with support for relating concepts, linking concepts to resources, merging ontologies, etc. The TM4L environment consists of an Editor and Viewer, available for download at <http://compsci.wssu.edu/iis/nsdl/download.html>.

The latest release of the TM4L Editor extends the Topic Map model with contexts thus allowing contextual retrieval of information that supports search of both relevant and similar resources. It also provides internationalization of the interface including translations in Spanish, traditional and simplified Chinese, French, German, Japanese, and Nepalese.

The TM4L development is funded by the NSF NSDL program. More information is available at <http://compsci.wssu.edu/iis/nsdl/index.html>

Related Resources

Dicheva, Darina, Christo Dichev, and Wang Dandan. [Visualizing Topic Maps for e-Learning](#). Fifth IEEE International Conference on Advanced Learning Technologies. DOI: [10.1109/ICALT.2005.305](https://doi.org/10.1109/ICALT.2005.305)

Dicheva, Darina, Christo Dichev. [Authoring educational topic maps: can we make it easier?](#). Fifth IEEE International Conference on Advanced Learning Technologies. DOI: [10.1109/ICALT.2005.73](https://doi.org/10.1109/ICALT.2005.73)

Dicheva, Darina, and Christo Dichev. 2006. [Topic Maps 4 E-Learning](#). <http://compsci.wssu.edu/iis/nsdl/>.

Dicheva Darina., Christo Dichev. 2006. [TM4L: Creating and Browsing Educational Topic Maps](#), *British Journal of Educational Technology - BJET*, 37(3), 391-404.