

Networked Knowledge Organization for the Networked Digital Library of Theses and Dissertations

**ECDL 2005, Vienna
4th European**

**Networked Knowledge Organization Systems
(NKOS) Workshop
September 22, 2005**

Edward A. Fox, Baoping Zhang, and
Ryan Richardson
Virginia Tech, Blacksburg, VA 24061 USA

fox@vt.edu

<http://fox.cs.vt.edu>

<http://www.ndltd.org>

Acknowledgements (Selected)

- **NDLTD Board of Directors, Committees**
- **Sponsors:** Adobe, FIPSE (US Dept. of Education), NSF (IIS-9986089, 0086227, 0080748, 0325579; DUE-0121679, 0136690, 0121741, 0333601), OCLC, UNESCO, VTLS
- **Faculty/Staff:** Tony Atkins, Debra Dudley, John Eaton, Gail McMillan, ...
- **VT (Former) Students:** Shahrooz Feizabadi, Marcos Goncalves, Ming Luo, Hussein Suleman, ...

Outline

- NDLTD Overview
- Key Challenges
- Concept Map Research
- Classification Research
- Selected Links

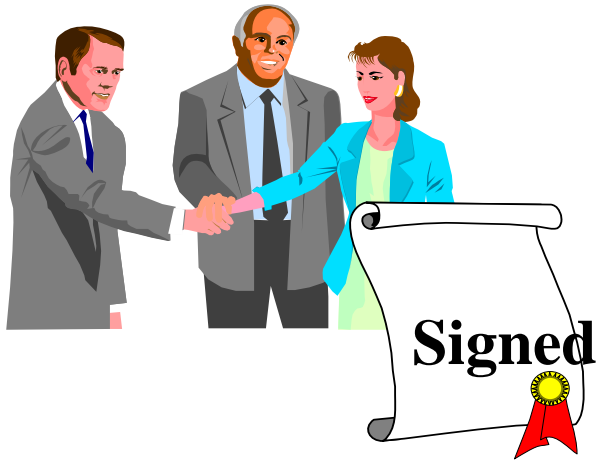
The Networked Digital Library of Theses and Dissertations

www.NDLTD.org

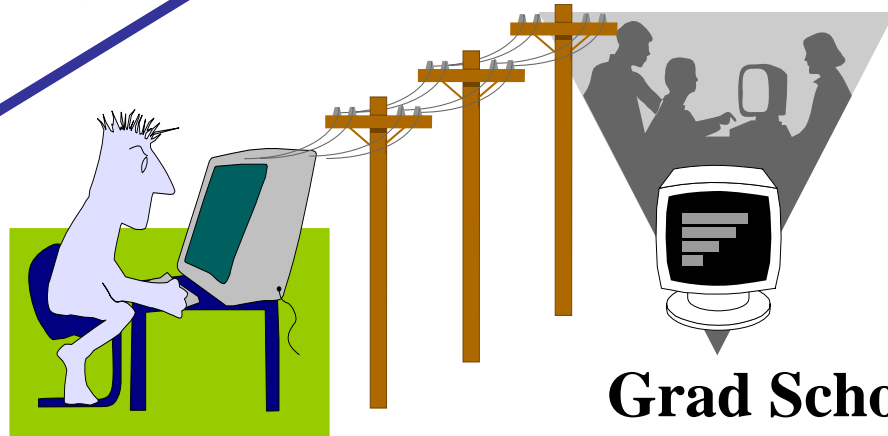
Training Authors
Expanding Access
Preserving Knowledge
Improving Graduate Education
Enhancing Scholarly Communication
Empowering Students & Universities

Leader of the Worldwide ETD
(Electronic Thesis and Dissertation) Initiative

Student Gets Committee Signatures and Submits ETD



Approval form



Library Catalogs ETD, Access is Opened to the New Research



Digital library access control

What are the long term goals?

- 400K US students / year getting grad degrees are exposed / involved
- 200K/yr rich hypermedia ETDs
- Possible evolution into electronic portfolios (data, images, video, audio, ...) for projects
- Dramatic increase in knowledge sharing: literature reviews, bibliographies, ...
- Services providing lifelong access for students: browse, search, prior searches, citation links
- Hundreds/thousands of downloads / yr / work

Why ETD? Short Answer

- **For Students:**
 - Gain knowledge and skills for the Information Age
 - Richer communication (digital information, multimedia, ...)
- **For Universities:** Easy way to enter the digital library field and benefit thereby
- **For the World:** Global digital library – large, useful, many services
- **General:**
 - Save time and money
 - Increased visibility for all associated with research results

ETD-MS

- ETD Metadata Standard
 - XML-encoded metadata standard (content and encoding) for Electronic Theses and Dissertations (ETDs)
 - in part conforming to Dublin Core (DC)
 - using RDF
 - using UNICODE
- Has specified relationship with MARC.

ETD Cataloging

- VT retained old cataloging policies, except:
 - author-assigned keywords (not LCSH)
 - generic (not LC) call no.
 - fields/subfields as req'd for computer files
 - full abstracts.
- USF: catalogers must do even better.
- General (LoC/Dewey) vs. for discipline:
 - OCLC has 3M records of TDs from local catalogs.
 - PhysNet: use PACS (VT study with training so can be semi-automatically?)

Workshop Key Challenges

- ***User-centered design issues***
- ***Mapping between different KOS***
- ***KOS representations and service protocols***
- VT has much related research, among our projects.

Cluster NDLTD-Computing

The screenshot shows a Microsoft Internet Explorer browser window with the address bar displaying `http://thorn.dlib.vt.edu:8080/Offcontroller/newsearch.do?processingChain=grapezone.process.scri`. The page content includes a search bar with the query `Rao: NDLTD, English stemmer, Tf Terms weighing, AHC, Dynamic Tree`. Below the search bar, there is a sidebar with a tree view of search results under the heading `sub topics`. The main content area displays two search results:

- 99 Assessing the impact of XML/EDI with real option valuation**

Electronic Data Interchange (EDI) hat seit über 25 Jahren das Ziel des nahtlosen elektronischen Austausch strukturierter Geschäftsdaten. Schätzungen nach nutzen jedoch lediglich 5 Prozent der Unternehmen EDI. Der Hauptgrund liegt darin, dass kleine und mittelständische Unternehmen die relativ hohen Setup- und Betriebskosten einer EDI-Lösung scheuen. Hinzu kommt die Unsicherheit bezüglich der EDI-Standardisierungsentscheidungen anderer Akteure. Durch die Verwendung von XML kann der Austausch von Informationen zwischen verschiedenen Anwendungen, wie z. B. bei EDI, sowohl in seiner Struktur als auch durch das Übertragungsmedium Internet stark vereinfacht und verbilligt werden und somit die Flexibilität von internet-basierten Geschäftsbeziehungen erhöhen. Der Standardisierungsprozess von XML/EDI ist bereits im Gang. Für Unternehmen stellt sich heute die Frage, ob sich eine Investition in XML/EDI rechnet. Hierzu wird das Black-Scholes Modell zur Bewertung von Real Optionen verwendet, wobei die Parameter der Formel in Bezug auf die XML/EDI Investitionsentscheidung näher analysiert werden, um den optimalen Investitionszeitpunkt zu bestimmen. (i) Projektkosten der Implementierung, (ii) Wert der Einsparungen durch das Projekt (substitutionale, komplementäre und strategische Einsparungen), sowie das (iii) Projektrisiko (interne und externe Risiken). Ausgehend von einer Analyse der technischen Eigenschaften von XML und EDI werden die Parameter der Formel Schrittweise in eine Entscheidungsmatrix eingebracht. Vier Firmenprofile, basierend auf der Theorie von Technologiediffusion, werden dargestellt um die Parameter der Black-Scholes Formel greifbarer darzustellen. (Autorenreferat)
http://epub.wu-wien.ac.at/dyn/dl/diss/epub-wu-01_1a2
- 100 Statistical data compression by optimal segmentation**

Die Arbeit behandelt die statistische Datenreduktion bzw. Datenkompression durch eine allgemeine Klasse von Klassifikationsmethoden. Die Datenkompression erzeugt eine Repräsentation des Datensatzes durch eine Partition oder durch charakteristische Punkte (genannt Prototypen). Diese Optimierungsprobleme stehen in Beziehung zu Partitionen mit minimaler Varianz und zum Principal-Point-Problem. Zur Lösung dieser Probleme wird eine Fixpunktmethode und ein adaptiver Ansatz verwendet. Die Arbeit enthält eine Darstellung der theoretischen Grundlagen des Optimierungsproblems und eine Reihe von Pseudo-Codes für die numerische Lösung der Datenkompression. Der wesentliche Teil behandelt praxisbezogene Fragestellungen zur Durchführung der Datenkompression. Dazu gehören die Bestimmung einer geeigneten Anzahl an charakteristischen Punkten, die Wahl einer Zielfunktion, die Implementierung einer Nachbarschaftsstruktur und die Verbesserung des Fixpunktalgorithmus. Die Eignung der vorgeschlagenen Methoden und Algorithmen wird anhand von Experimenten vorgeführt und durch eine Reihe von Beispielen illustriert. (Autorenreferat)
http://epub.wu-wien.ac.at/dyn/dl/diss/epub-wu-01_26b

The browser window also shows a sidebar with a tree view of search results under the heading `sub topics`. The tree view includes categories such as `Design (45)`, `Implementation, Performance (23)`, `User (14)`, and `Development, Software (18)`.

Example of Union Service: CitiViz

CitiViz

Data Compression Search

Search for ACM Only Search for all collections

X-axis attributes: Citation Y-axis attributes: Rank

Show all data in the scatter plot Reset the scatter plot

- ▶ D.4.3: File Systems Management (Del)
- ▶ D.4.7: Organization and Design (Del)
- ▶ **E 1: Data Structures** (Del)
 - Data compression with finite windows
 - Application of splay trees to data compression
 - Practical prefetching via data compression
- ▶ E 2: Data Storage Representations (Del)
- ▶ E 4: Coding And Information Theory (Del)
- ▶ E 5: Files (Del)
- ▶ F.1.3: Complexity Measures and Classes (Del)
- ▶ F.2.2: Nonnumerical Algorithms and Problems (Del)
- ▶ F.3.2: Semantics of Programming Languages (Del)
- ▶ G.2.2: Graph Theory (Del)
- ▶ H.2.1: Logical Design (Del)
- ▶ H.2.7: Database Administration (Del)
- ▶ H.2.8: Database Applications (Del)
- ▶ H.3: Information Storage And Retrieval (Del)

Previous Next

Rank	15
Title	Application of splay trees to data compression
Author	D. W. Jones
Published date	1988
Collection	ACMDL
Abstract	The splay-prefix algorithm is one of the simplest and fastest adaptive data compression algo...
url	http://www.citidel.org/?op=getobj&identifier=oai:ACMDL:articles.63036
Citation	12

H.3: Information Storage And Retrieval
E.4: Coding And Information Theory
D.4.3: File Systems Management

CC

CCS

CORR

msc

CCS1998 :: Computing Milieux

[Computers And Education](#) (7164)[Computers And Society](#) (3038)[General](#) (163)[History Of Computing](#) (597)[Legal Aspects Of Computing](#) (479)[Management Of Computing And Information Systems](#) (7191)[Miscellaneous](#) (7)[Personal Computing](#) (406)[The Computer Industry](#) (1050)[The Computing Profession](#) (1734)

Tip: Not all resources are categorized. If you can't find what you need here, try issuing a whole-site search above.

Search in this category: Showing 1 - 10 of 18156 resources in this category: (16 resources filtered out. [show all.](#))page: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [11](#) [12](#) [13](#) [14](#) [15](#) >> >| of 1816 (18156 items)jump directly to page #:

1. [Reusability and adaptability of interactive resources in web-based educational](#)

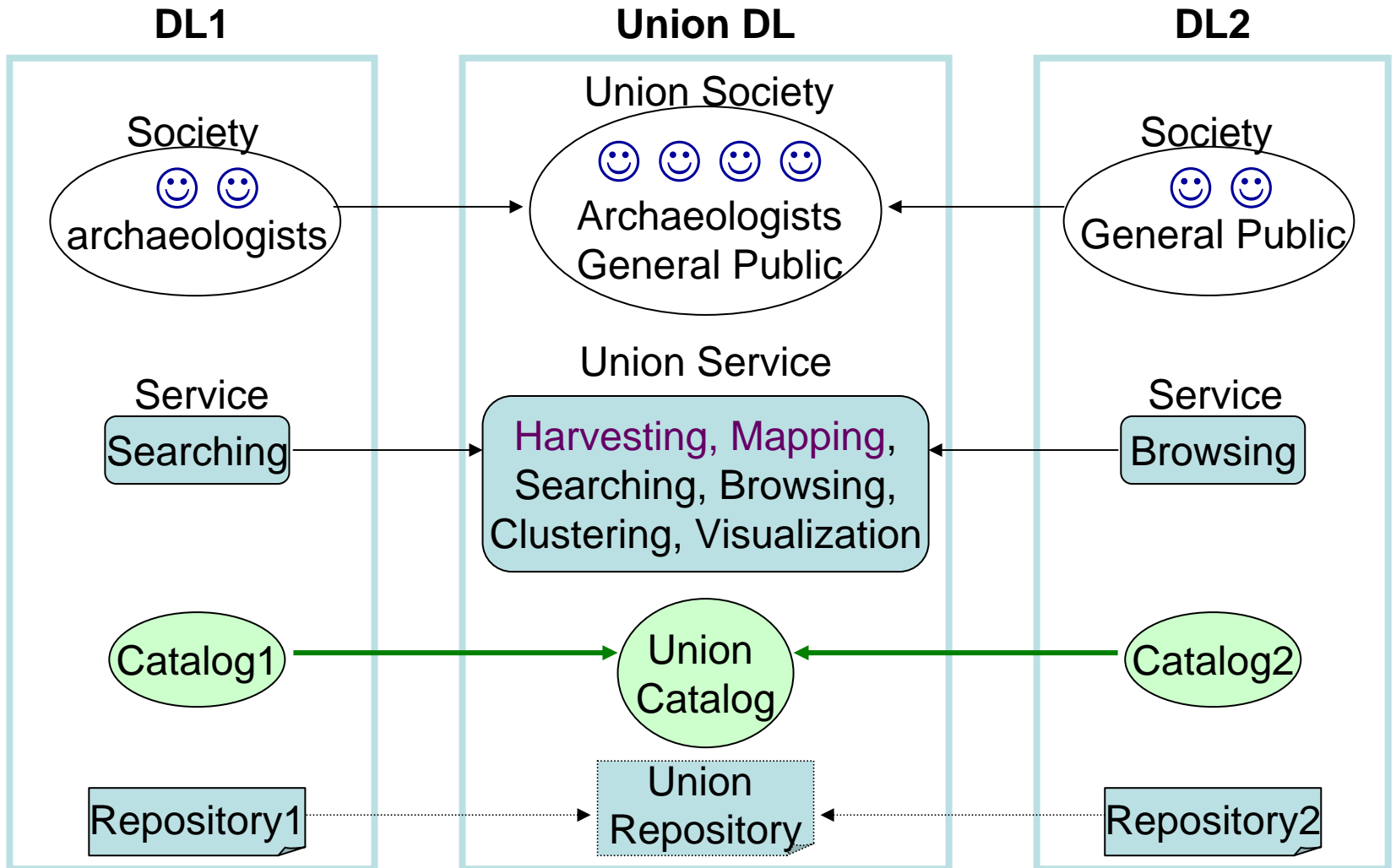
The production of interactive multimedia content is in most cases an expensive task in terms of time and cost. Hence, optimizing production by exploiting the reusability of interactive multimedia elements is mandatory. Reusability can be triggered by a combination of reusable multimedia components and the appropriate use of metadata to control the components as well as their combination. In this article, we discuss the reusability aspects of interactive multimedia content in web-based learning systems. In...

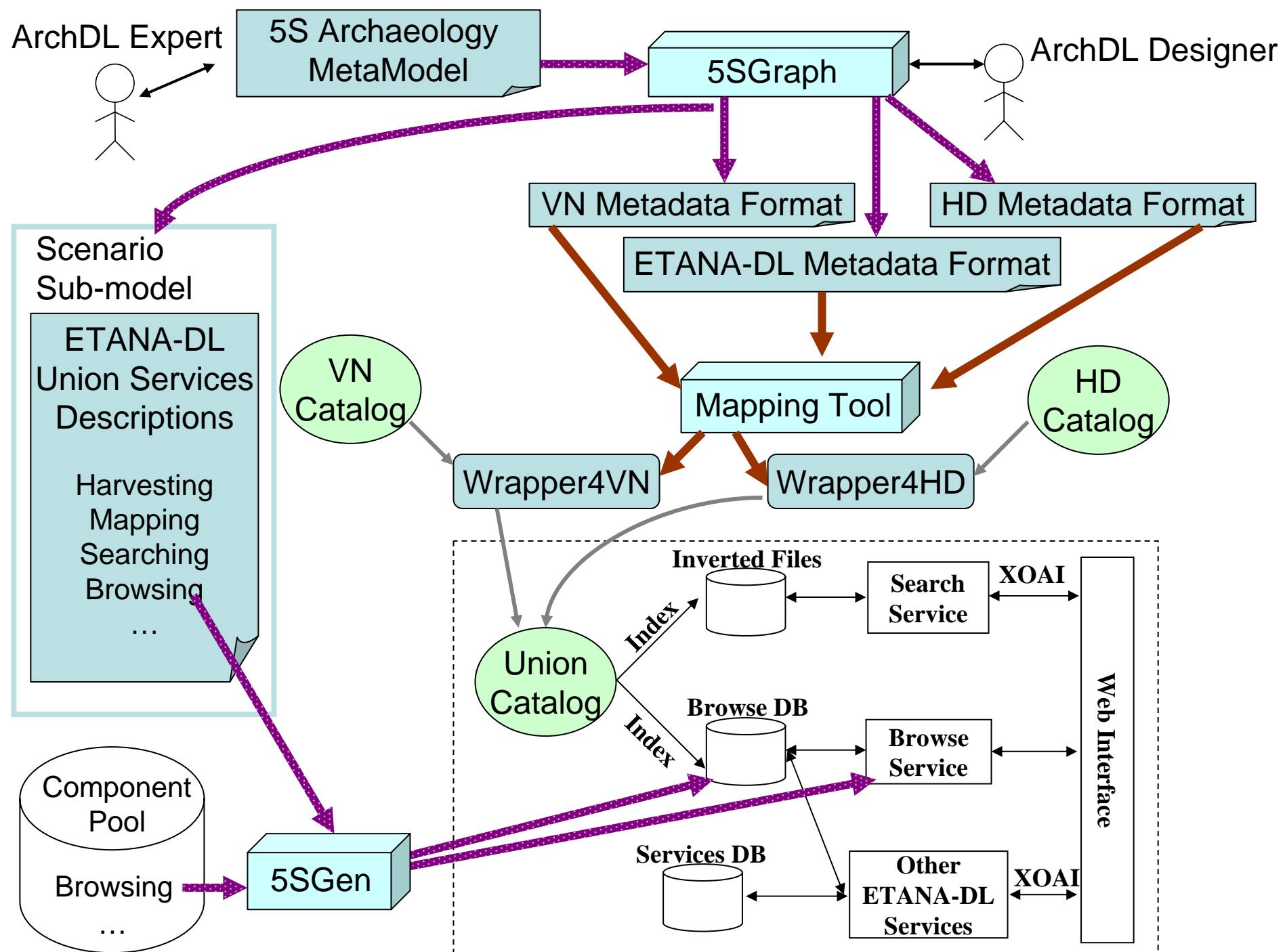
[By: Abdulmotaleb El Saddik, Stephan Fischer, Ralf Steinmetz] [Date: 2001-08-21] [From: [CSTC](#)]

2. [Altair Emulator](#)

This is a very well done emulator of the Altair 8800b

Architecture of a Union DL





Concept Map Research

- Research topics:
 - Using concept maps as a summarization technique for ETDs
 - Translating concept maps to allow for cross-language relevance determination
- We found that taking advantage of structure of documents (table of contents, etc) can improve the automatically generated maps.

Concept Map Research (cont'd)

We are looking for

- tools for automatic phrase extraction and translation using ETDs as the corpus
 - something better than WordNet to determine relationships between concepts, and a
 - collection of ontologies to cover the scope of ETDs.
-
- Can these ontologies be combined to cover broader scopes of knowledge?

Classification Research

- How to identify the language of an ETD?
- How to identify the subject discipline(s) of an ETD?
 - This could support browsing by categories.
 - For some ETDs, we may have the degree discipline field in the ETD-MS metadata.
 - But for most of them, we lack this information.
 - If training is used, what data would be of greatest use, to cover all disciplines?

Classification Research (cont'd)

- We could use machine learning methods to combine different evidence to automatically classify works. What is best?
- What kinds of evidence may be used to help automatic text classification?
 - structural content?
 - citations?
 - classification scheme hierarchy of the (tree structure)?

Summary

- NDLTD Overview
- Key Challenges
- Concept Map Research
- Classification Research
- Selected Links

Selected Links

- Fox: <http://fox.cs.vt.edu>
- CITIDEL (computing education resources)
 - www.citidel.org
- NDLTD (electronic theses and dissertations worldwide)
 - www.ndltd.org, etdguide.org
- OAI (Open Archives Initiative)
 - www.openarchives.org
- Virginia Tech Digital Library Research Laboratory (DLRL, www.dlib.vt.edu)
 - 5S, ENVISION, MARIAN, ...)