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

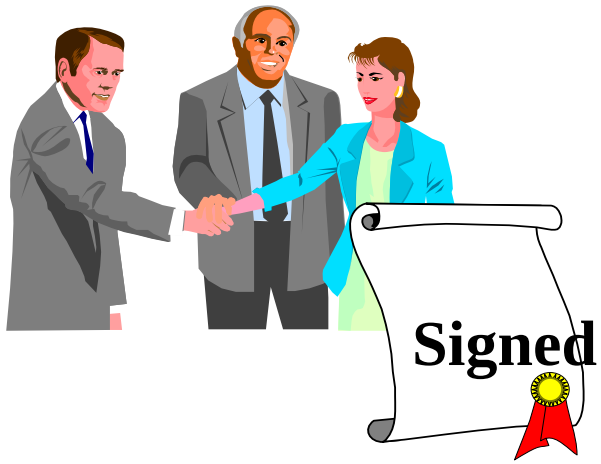
The Networked Digital Library of Theses and Dissertations

www.NDLTD.org

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

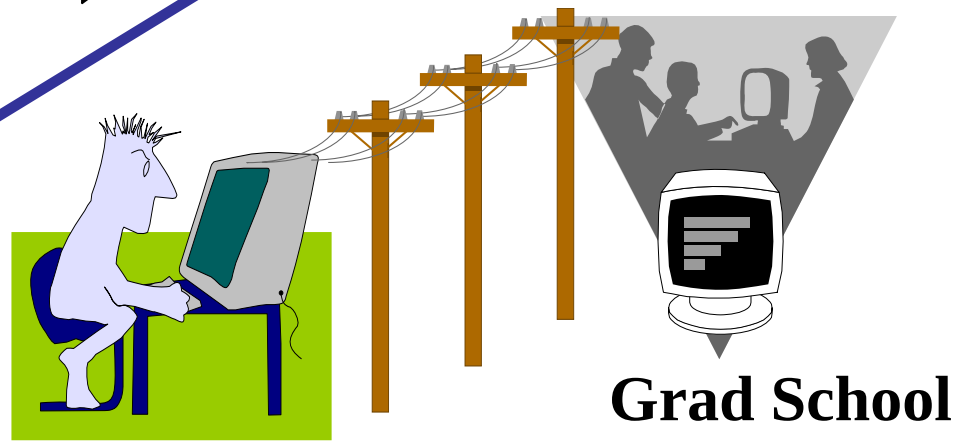
Leader of the Worldwide ETD
(Electronic Thesis and Dissertation) Initiative

Student Gets Committee Signatures and Submits ETD



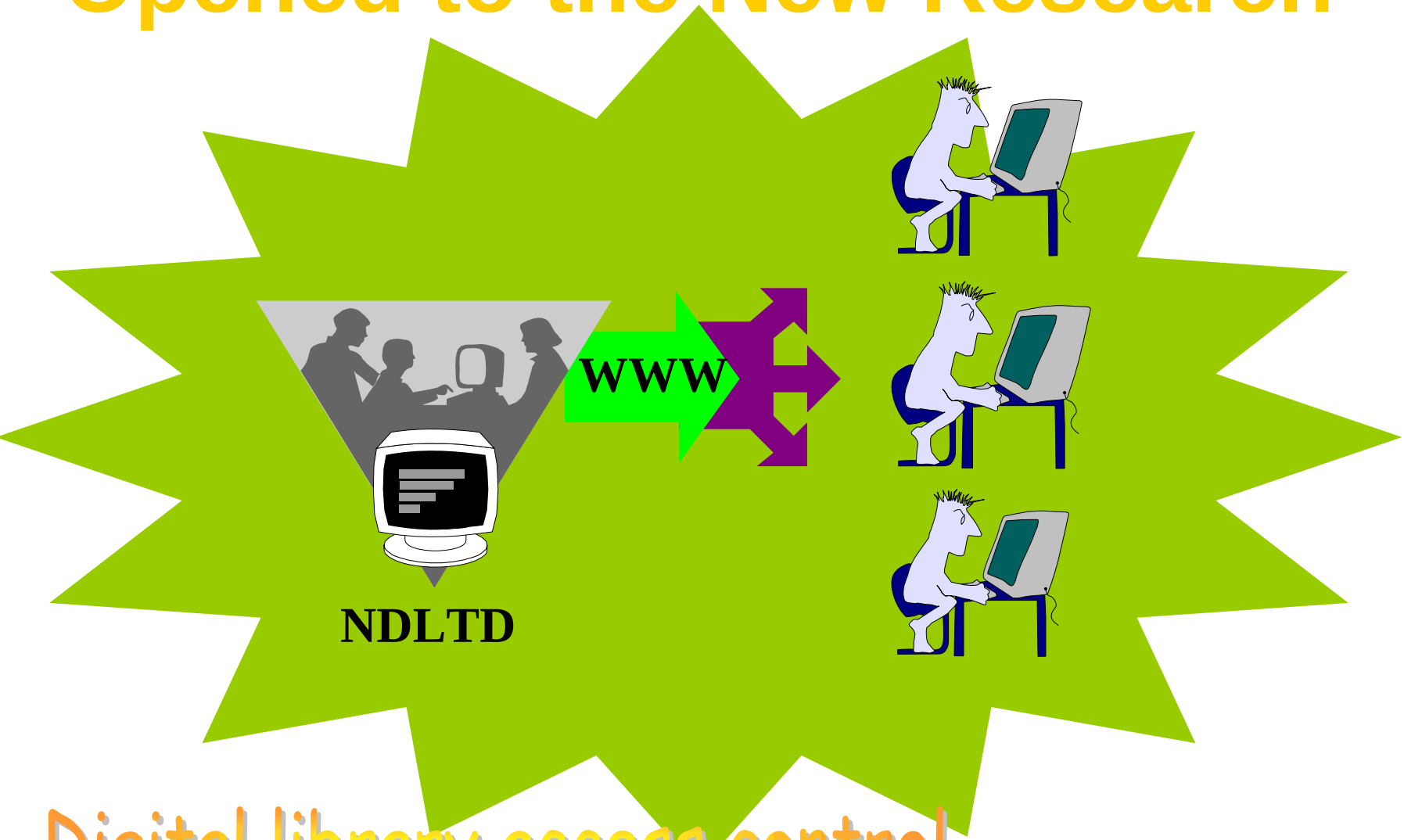
Signed

Approval form



Grad School

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

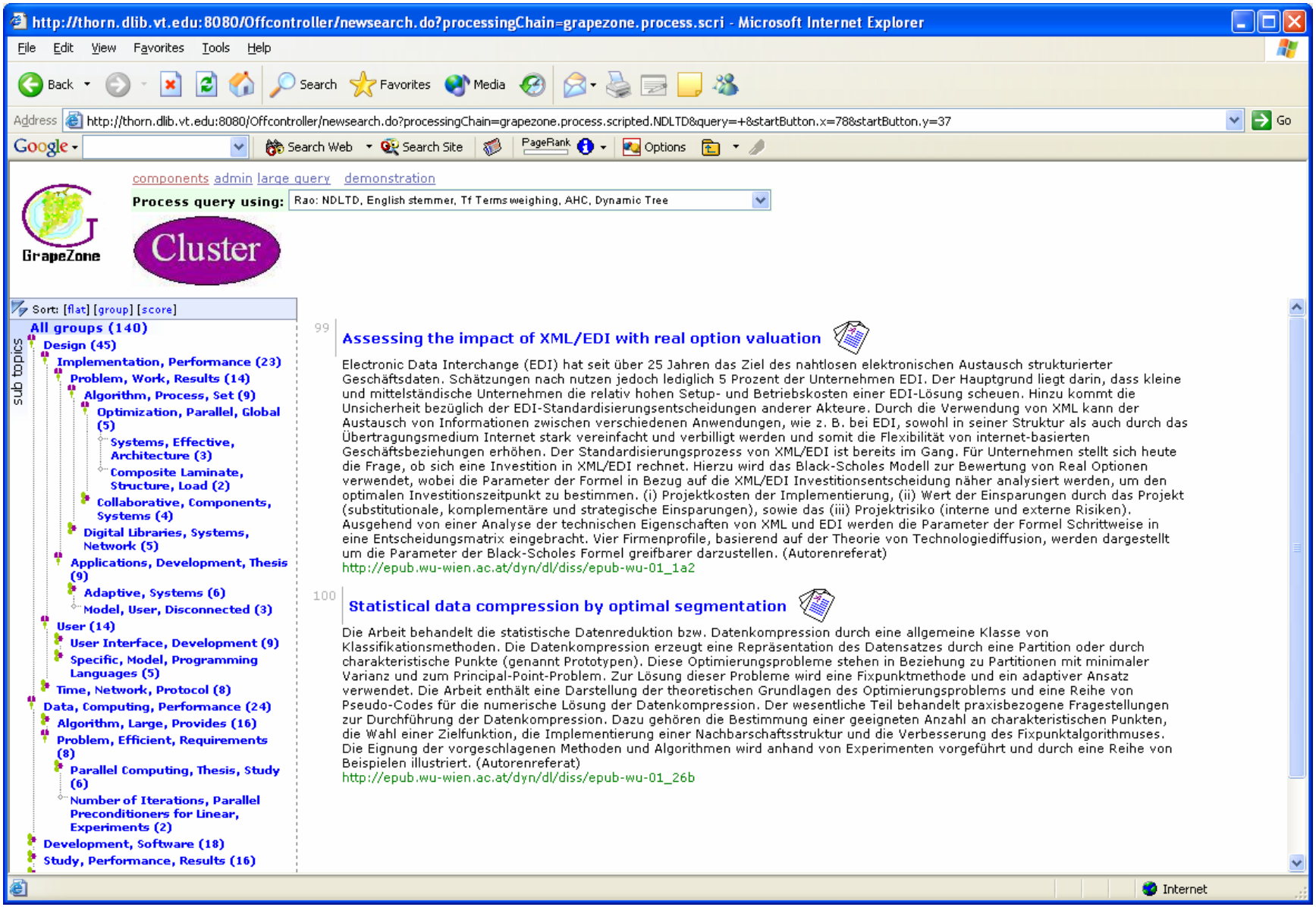
- VT retained old cataloging policies, except:
  - author-assigned keywords (not LCSH)
  - generic (not LC) call no.
  - fields/subfields as req'd for computer files
  - full abstracts.
- USF: catalogers must do even better.
- General (LoC/Dewey) vs. for discipline:
  - OCLC has 3M records of TDs from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

Workshop Key Challenges

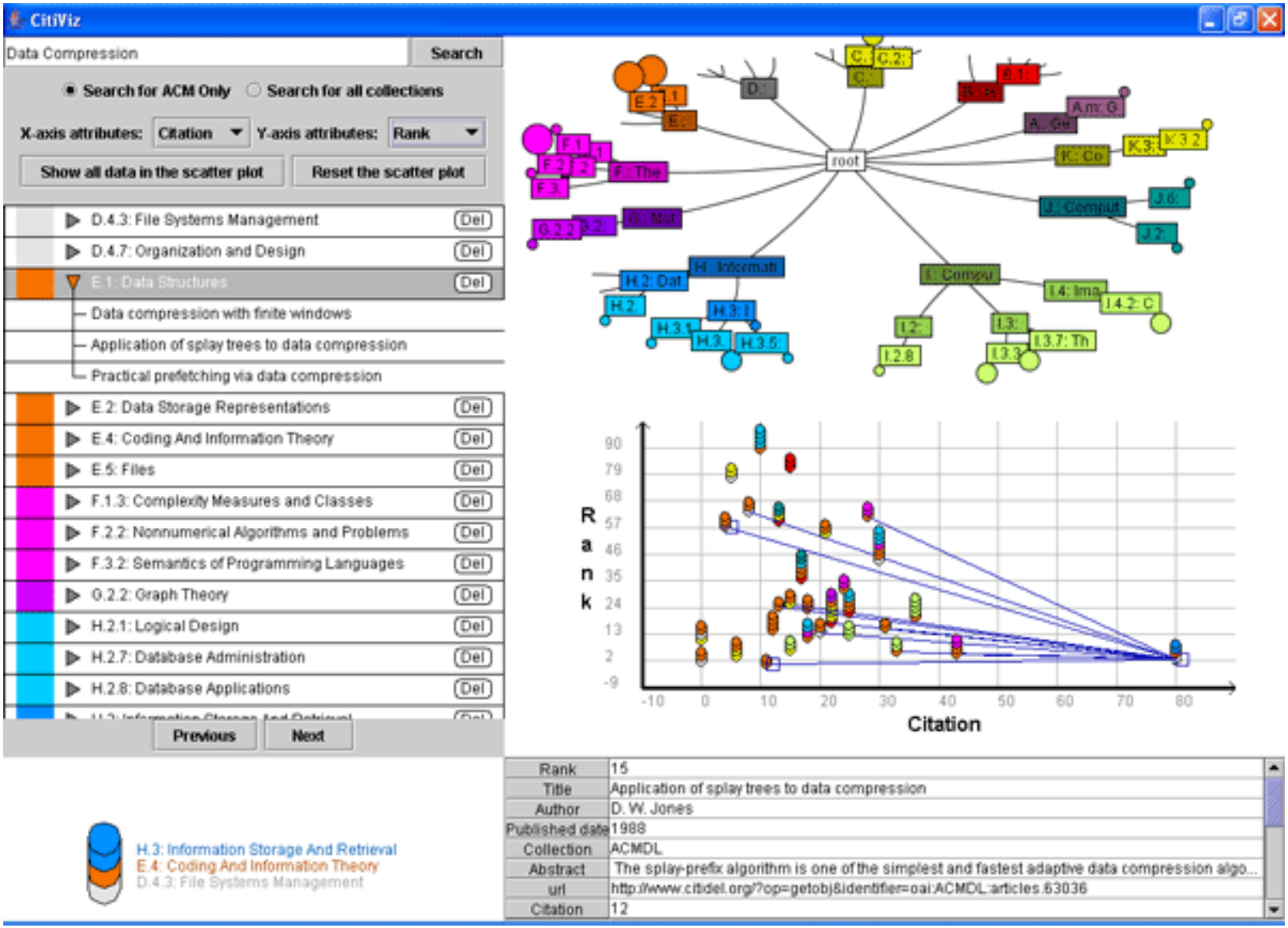
- ***User-centered design issues***
- ***Mapping between different KOS***
- ***KOS representations and service protocols***

- VT has much related research, among our projects.

Cluster NDLTD-Computing



Example of Union Service: CitiViz



CCS1998 :: Computing Milieux

Computers And Education *(7164)*	Management Of Computing And Information Systems *(7191)*
Computers And Society *(3038)*	Miscellaneous *(7)*
General *(163)*	Personal Computing *(406)*
History Of Computing *(597)*	The Computer Industry *(1050)*
Legal Aspects Of Computing *(479)*	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: go!

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 # : go

1. Reusability and adaptability of interactive resources in web-based educational

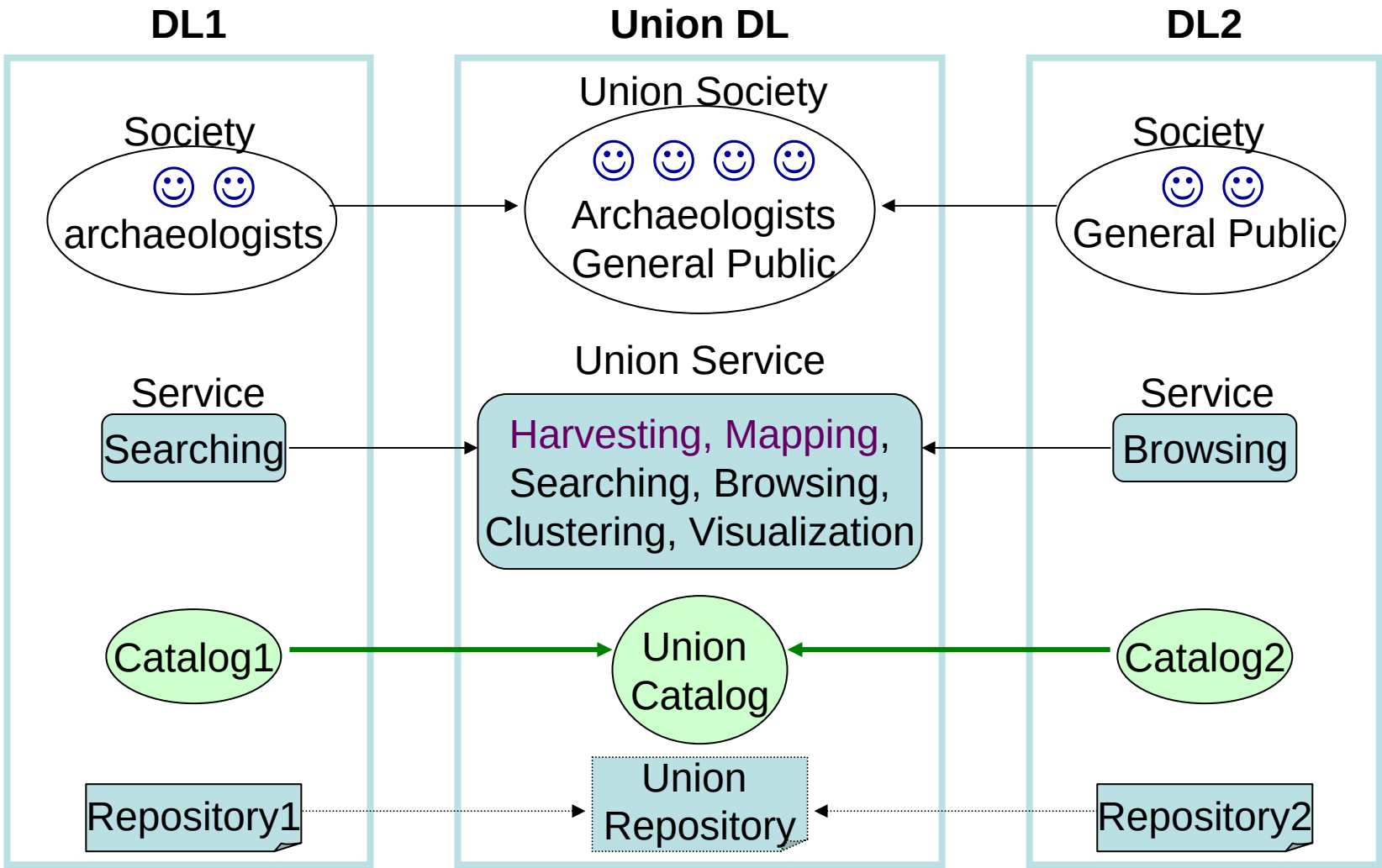
The production of interactive multimedia content is in most cases an expensive task in terms of time and cost. Hence, optimizing production by exploiting the reusability of interactive multimedia elements is mandatory. Reusability can be triggered by a combination of resuable multimedia components and 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...

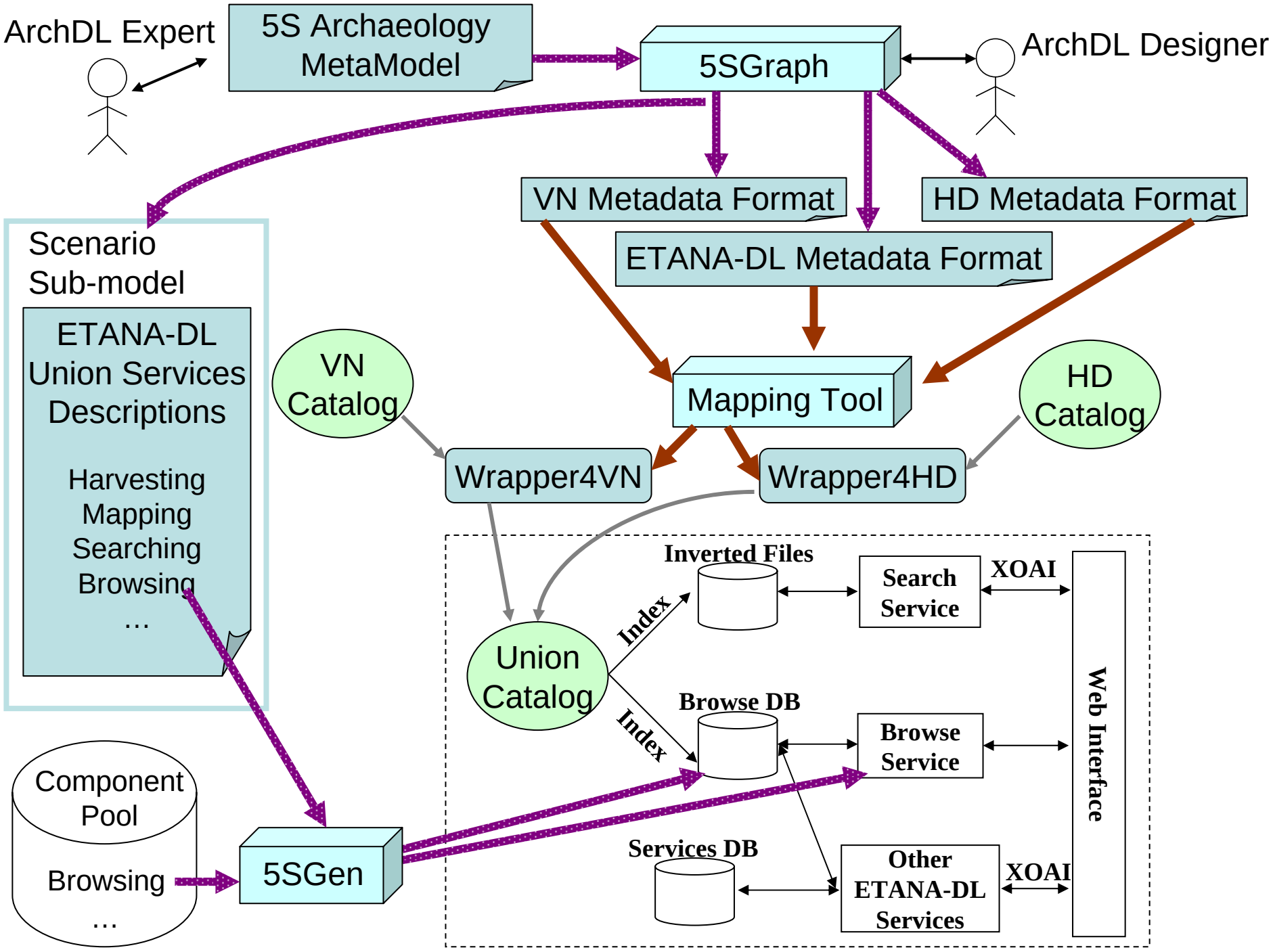
[By: Abdulmotaleb El Saddik, Stephan Fischer, Ralf Steinmetz] [Date: 2001-08-21] [From: CSTC]

2. Altair Emulator

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

Architecture of a Union DL




ArchDL Expert
5S Archaeology MetaModel
5SGraph
ArchDL Designer
VN Metadata Format
HD Metadata Format
ETANA-DL Metadata Format
Scenario Sub-model
ETANA-DL Union Services Descriptions
Harvesting
Mapping
Searching
Browsing
...
VN Catalog
Mapping Tool
HD Catalog
Wrapper4VN
Wrapper4HD
Inverted Files
Search Service
XOAI
Index
Union Catalog
Index
Browse DB
Browse Service
Web Interface
Component Pool
Browsing
...
5SGen
Services DB
Other ETANA-DL Services
XOAI

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, …)