IBM Academic Initiative Introduction for Pamplin School of Business

Virginia Tech – October 13, 2011

“IBM Academic Skills Cloud and Computing Education Modules”

by Edward A. Fox

• fox@vt.edu    http://fox.cs.vt.edu
• Dept. of Computer Science, Virginia Tech
• Blacksburg, VA 24061 USA
Background: Digital Libraries (DL) Curriculum Project

- NSF awards to VT and UNC-CH
  - Grant Numbers IIS-0535057 & 0535060
- CS and LIS
- Project server: http://curric.dlib.vt.edu/
- Coverage: Digital libraries, but extended to Information retrieval, Multimedia
## Software-based Modules and Cloud Instances

<table>
<thead>
<tr>
<th>IR</th>
<th>Apache Solr</th>
<th>Lemur</th>
</tr>
</thead>
<tbody>
<tr>
<td>IR</td>
<td>WordNet</td>
<td>R</td>
</tr>
<tr>
<td>IR</td>
<td>NLTK</td>
<td>SEDNA XML DB</td>
</tr>
<tr>
<td>IR</td>
<td>CLUTO</td>
<td>Weka</td>
</tr>
<tr>
<td>IR</td>
<td>TREC Eval</td>
<td>Hadoop Map-Reduce</td>
</tr>
<tr>
<td>MM</td>
<td>Media Computation</td>
<td>Audacity</td>
</tr>
<tr>
<td>MM</td>
<td>PureData</td>
<td>Fingerprint</td>
</tr>
</tbody>
</table>
Cloud Module Development

• Fall 2010 – CS5604 (Information Retrieval)
  – 5 cloud instances developed, then used
  – 5 additional cloud instances developed

• Spring 2011 – CS4624 (Multimedia, Hypertext, and Information Access)
  – Use of above cloud instances
  – 4 additional cloud instances developed

• Fall 2011 – CS6604 (Digital Libraries)
  – Update 8 of the cloud instances
  – Update 6 of the other modules
Module-based Pedagogy

- Class use of modules, 1 wk each
- Independent study of a module of interest
- Independent study prepping for tool use

- Learning by making modules
- Learning by preparing cloud instances
Module Template

1. Module name
2. Scope
3. Learning objectives
4. 5S characteristics of the module
   (streams, structures, spaces, scenarios, societies)
5. Level of effort required (in-class and out-of-class time required for students)
6. Relationships with other modules (flow between modules)
7. Prerequisite knowledge/skills required
8. Introductory remedial instruction
9. Body of knowledge (theory + practice;
an outline that could be used as the basis for class lectures)
10. Resources (required readings for students;
    additional suggested readings for instructor and students)
11. Exercises / Learning activities
12. Evaluation of learning objective achievement
    (graded exercises or assignments)
13. Glossary
14. Additional useful links
15. Contributors (authors of module, reviewers of module)
Completed Topical Modules - 1

- Module 1-a (10-c): Conceptual frameworks, models, theories, defns
- Module 1-b: History of digital libraries and library automation
- Module 2-c (8-d): File Formats, Transformation, and Migration
- Module 3-b: Digitization
- Module 3-e (7-e): Web publishing
- Module 4-b: Metadata
- Module 5-a: Architecture overviews
Completed Topical Modules - 2

- Module 5-b: Application software
- Module 5-d: Protocols
- Module 6-a: Information needs/relevance
- Module 6-b: Online information seeking behaviors and search strategies
- Module 6-d: Interaction design and usability assessment
- Module 7-a: Indexing and searching
- Module 7-a(1): Image retrieval
Completed Topical Modules - 3

- Module 7-b: Reference Services
- Module 7-c: Recommender systems
- Module 7-d: Routing
- Module 7-f: Crawling
- Module 7-g: Personalization
- Module 8-a: Preservation
- Module 8-b: Web archiving
- Module 9-c: Digital library evaluation, user studies
### FRAMEWORK FOR A DIGITAL LIBRARY CURRICULUM

**(2008/08/23)**

<table>
<thead>
<tr>
<th>CORE TOPICS</th>
<th>1-a (10-c): Conceptual frameworks, models, theories, definitions</th>
<th>1-b: History of digital libraries and library automation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Overview</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Digital Objects</td>
<td>2-a: Text resources</td>
<td>2-c (8-d): File formats, transformation, migration</td>
</tr>
<tr>
<td>3 Collection Development</td>
<td>3-a: Collection development/selection policies</td>
<td>3-e (7-e): Web (push) Publishing</td>
</tr>
<tr>
<td>4 Info/ Knowledge Organization</td>
<td>4-a: Information architecture (e.g., hypertext, hypermedia)</td>
<td>4-d: Subject description, vocabulary control, thesauri, terminologies</td>
</tr>
<tr>
<td>5 Architecture (agents, mediators)</td>
<td>5-a: Architecture overviews</td>
<td>4-e: Object description and organization for a specific domain</td>
</tr>
<tr>
<td>6 User Behavior/Interactions</td>
<td>6-a: Info needs, relevance</td>
<td>5-d: Protocols</td>
</tr>
<tr>
<td>7 Services</td>
<td>7-a: Indexing and searching</td>
<td>5-e: Interoperability</td>
</tr>
<tr>
<td>8 Preservation</td>
<td>8-a: Preservation</td>
<td>5-f: Security</td>
</tr>
<tr>
<td>9 Management and Evaluation</td>
<td>9-a: Project management</td>
<td>6-c: Sharing, networking, interchange (e.g., social)</td>
</tr>
<tr>
<td>10 DL education and research</td>
<td>10-a: Future of DLs</td>
<td>6-d: Interaction design, usability assessment</td>
</tr>
<tr>
<td></td>
<td>10-b: Education for digital librarians</td>
<td>6-e: Info summarization and visualization</td>
</tr>
</tbody>
</table>

---

1 This project is funded by the National Science Foundation through grants NSF IIS-0535057 (to Virginia Tech) and IIS-0535060 (to University of North Carolina at Chapel Hill).
DL Curriculum Framework

Semester 1:
- DL collections: development/creation

Semester 2:
- DL services and sustainability

COURSE STRUCTURE

Digitization
Storage
Interchange

Metada Cataloging
Author submission

Architectures (agents, buses, wrappers/mediators)
Interoperability

Naming Repositories
Archives

Archiving and preservation
Integrity

CORE DL TOPICS

Digital objects
Composites
Packages

Spaces (conceptual, geographic, 2/3D, VR)

Architectures (agents, buses, wrappers/mediators)
Interoperability

Intellectual property rights mgmt.
Privacy Protection (watermarking)

Documents
E-publishing
Markup

Multimedia streams/structures
Capture/representation
Compression/coding

Thesauri
Ontologies
Classification
Categorization

INFO. Needs
Relevance
Evaluation
Effectiveness

Bibliographic information
Bibliometrics
Citations

Content-based analysis
Multimedia indexing

Multimedia presentation, rendering

Routing Filtering
Community filtering

Search & search strategy
Info seeking behavior
User modeling
Feedback

Info summarization
Visualization
<table>
<thead>
<tr>
<th>CC2001 Information Management Areas</th>
</tr>
</thead>
<tbody>
<tr>
<td>IM1. Information models and systems*</td>
</tr>
<tr>
<td>IM2. Database systems*</td>
</tr>
<tr>
<td>IM3. Data modeling*</td>
</tr>
<tr>
<td>IM4. Relational DBs</td>
</tr>
<tr>
<td>IM5. Database query languages</td>
</tr>
<tr>
<td>IM6. Relational DB design</td>
</tr>
</tbody>
</table>

* Core components
Questions?
Discussion?

Thank You!
(fox@vt.edu)