SI 639: An Experiment in Teaching Web Archiving

Margaret Hedstrom
Motivation/Context

• New Master’s Specialization “Preserving Information” (PI)
• Emphasis on Digital Preservation
• Market Demand for Graduates with Web Archiving Skills
• Frustration with “Manual” Web Archiving
Format/Approach

• Half-term course (7 weeks) 1.5 credits
• Lecture (90 minutes)/Lab (90 minutes)
• Lab Assignments (most weeks) leading to a Final Project (produce a web archive on topic of your choice)
Course Objectives

1. Become familiar with the challenges of acquiring, downloading, storing, and providing access to web-based content.

2. Learn how to program and use web archiving tools.

3. Learn how to create a focused archive of web content.

4. Become aware of legal and policy constraints to web archiving.

5. Learn standards and best practices for sustainability of archived web content.
Course Content

- Basic Concepts and Tools for Web Archiving
- Selection of Web Content
- Acquiring Web Content for Archiving (using HTTrack Web Archiving Tool)
- Archiving Deep Web Content
- Quality Control and Post Capture Processing
- Preserving Web Sites
- Access and Use of Web Archives
Issues

• Textbook
• Difficult to get sufficient practice beyond programming and crawling
• Technical support for a variety of web archiving tools
• Individual attention to problems/issues
• Depth of experience
Conclusions

• Lab and hands-on experience were positive
• Course will be offered again in Winter 2010
• Increase practical experience
• Add more tools
• Place more emphasis on preservation