LIS 7911 Construction of Index Languages and Thesauri
School of Library and Information Science, Louisiana State University
Spring 2011

Instructor:
Yejun Wu, PhD
Office: 272 Coates Hall, LSU
Tel: (225)578-1489
Email: wuyj@lsu.edu
http://www.csc.lsu.edu/~wuyj
Office hour: Tuesday 11AM-12PM
Class Time: Tuesday 5:45-8:30PM
Classroom: 263A Coates Hall

Prerequisite: LIS 7002 or LIS 7005 or LIS 7606 or LIS 7607 or permission of instructor

CATALOG DESCRIPTION
The designing and construction of index languages/thesauri. Analysis and evaluation of existing index languages/thesauri. Term project in constructing an index language/thesaurus.

COURSE OBJECTIVES
After completion of the course, the student
- Should be able to design an index language or thesaurus and construct it using manual methods with computer assistance.
- Should have a good understanding of the problems involved in and the procedures used for updating index languages and thesauri.
- Should have an understanding of the role of thesauri and ontologies in organizing materials on the Web.
- Should have some idea about automated and semi-automated methods in the construction of index languages and thesauri.

TEXTS
- Aitchison, Jean, Alan Gilchrist, and David Bawden. 2000. *Thesaurus construction and use: A practical manual*. 4th ed. London: ASLIB/IMI.
- Chapters of the following books will be assigned
  -- Soergel, D. (1974). Indexing Languages & Thesaurus: Construction & Maintenance. [need to get his permission to re-print some chapters]
  -- Soergel, D. (1985). Organizing Information – Principles of Database and Retrieval Systems.
  -- Cleveland, Donald B., and Ana D. Cleveland (2001). Introduction to indexing and abstracting. 3rd ed. Englewood, CO: Libraries Unlimited.

GRADE GUIDELINES
Course grade will be based on class participation, homework, main project (including project report), and final exam. They are weighted as follows:

- Class participation (including in-class exercises): 20%
- Homework: 20%
- Main project: 40%
- Final exam: 20%

ACADEMIC INTEGRITY
All students and the instructor are responsible for observing the highest standards of intellectual and personal honesty in every aspect of their careers at Louisiana State University. The penalties for academic dishonesty are severe and ignorance is not an acceptable defense.

COURSE DESIGN AND TEACHING METHODOLOGY
The class sessions will consist of lectures by the instructor, class discussions, in-class exercises, project and student presentations. Students are required to come to class having read the assigned readings for the week. Supplemental readings may be added.

COURSE CALENDAR

Week 1. Course overview
- Objectives, prerequisites, learning/teaching methods
- Description of mini-project, main project
- Discussion of forming groups for main project: select a subject
- Knowledge organization systems (KOS)
- Thesauri, taxonomies, ontologies and the semantic Web
- Thesaurus construction: general framework and overall organization
- Find thesauri: thesauri on the Internet
- Thesaurus software: TheW32

Homework

Project
- Forming groups and select a subject

Readings
- Buckland, M. (1999). "Vocabulary as a central concept in library and information science," available at http://people.ischool.berkeley.edu/~buckland/colisvoc.htm
- Fast, Karl, Fred Liese, and Mike Stekel (2002). "What is a controlled vocabulary?" http://www.boxesandarrows.com/view/what_is_a_controlled_vocabulary_
- Aitchison, J. & Clake, S. D. (2004). The thesaurus: A historical viewpoint, with a look to the future. *Cataloging & Classification Quarterly*, 37(3/4): 5-21.
- Rowley, Jennifer, "The Controlled Versus Natural Indexing Languages Debate Revisited: A Perspective on Information Retrieval Practice and Research." *Journal of Documentation*, 1994: 108-118.
- Thomas, A. R. (2004). Teach yourself thesaurus: Exercises, readings, resources. *Cataloging & Classification Quarterly*, 37(3/4): 23-34.

- Milstead, Jessica L. Use of thesauri in the full-text environment. http://www.bayside-indexing.com/Milstead/useof.htm

Week 2. Indexing
- Controlled vs. free vocabularies;
- Indexing languages

In–class exercises: image indexing

Homework
- Option 1: Nonfiction book indexing. Suggest general principles that could govern decision-making both in this indexing task and in indexing in general.
- Option 2: ERIC Indexing. Indexing two articles and compare with the indexer's results. Write a report.

Project
- Forming groups and select a subject

Readings:
- Cleveland, Donald B., and Ana D. Cleveland. 2001. Introduction to indexing and abstracting. 3rd ed. Englewood, CO: Libraries Unlimited. Chapter 4 (p. 35-40), Chapter 5 (p.48-55, 61-73), Chapter 6
- McRae, Linda. (2000). Indexing images for subject access: controlled vocabularies in the VISION Project. *Art Documentation*, 19: 4-9.
- Shatford Layne, Sara. (1994). Some issues in indexing of images. *Journal of the American Society for Information Science,* 45(8): 583-588.
- Berinstein, P. (1999). "The Big Picture – do you see what I see? Image indexing principles for the rest of us," *Online* 23(2): 85-86, 99.
- Grund, A.(1993). ICONCLASS: On subject analysis of iconographic representations of works of art. *Knowledge Organization*, 20: 20-29.
- Woll, Johanna. (2005). User access to digital image collections of cultural heritage materials: the thesaurus as pass-key. *Art Documentation*, 24(2): 19-28.

Week 3. The entity-relationship approach to thesaurus construction
- Functions of a thesaurus, classification, ontological knowledge base
- Classification for user support and learning
- Procedure of thesaurus construction

In-class exercises:
- Semantic factoring
- Building a hierarchy of elemental concepts

Homework

■ Mini-project

Project
■ Final formation of groups
■ Define thesaurus scenario and scope and compile a list of resources (step F1.1). The list should contain an example for each type of source. (Some sources may require interlibrary loan, therefore it is necessary to start step F1.2 now even though the mini-project is not finished yet.)

Readings:
■ E-R approach, in Soergel, D. (1985). *Organizing Information – Principles of Database and Retrieval Systems*. Chapter 3 and Section 9.1
■ The procedure of thesaurus construction: Soergel, D. (1974). *Indexing Languages & Thesaurus: Construction & Maintenance*. Sections F0.1 (including Fig. 52, 53), F0.4.3, F0.5.1, F0.6, F0.7.2, F1-F5.9; E0, E1.0, E1.1, E1.5, E1.6, E1.8 (skip sections marked "advanced")
■ Nielsen, Marianne Lykke (2006). Thesaurus Construction: Key Issues and Selected Readings, *Cataloging & Classification Quarterly*, 37: 3, 57 -74
■ Shearer, James R.(2006). A Practical Exercise in Building a Thesaurus. *Cataloging & Classification Quarterly*, 37: 3, 35 –56
■ Neelameghan, A. (2009). Building a thesaurus for a specialized subject: A case report. *Information Studies*, 15(1): 61-64.

Week 4. Thesaurus construction (facets)
■ Hierarchy from facet combination

In-class exercises:
■ Hierarchy from facet-combination
■ Facet analysis exercise

Homework
■ Mini-project (continue)

Project
■ Open discussion.

Readings:
■ Cleveland & Cleveland – Chapter 4 (p.40-47, 91-96)
■ Aitchison et al. A-C (p.1-16); F2 (p.68-78); E (37-48)
■ Conceptual structure: Soergel (1974) book Sections B4.3, B6, C1.3.1, C1.3.2, C1.4, C1.5, C1.6, C3, C4.

Week 5. Thesaurus construction (relationships)

In-class exercise
- Semantic relationship analysis

Project
- Discussion of project (on conceptual structure, procedure, collecting terms)
- Start collecting terms (step F1.2.2): 200 terms per member

Readings
- Aitchison et al. F-F1 (p.49-68)
- Soergel (1974) book. Sections C5, C6, C7
- Readings on thesaurus software and workflow
- Mazzocchi, F. et al. (2007). Relational semantics in thesauri: some remarks at theoretical and practical levels. *Knowledge Organization*, 4. 197-246

Week 6. Thesaurus construction (final term selection)

Project
- Group discussion: scenario and scope, list of sources, methodology of transferring terms and relationships
- Finish collecting and inputting terms (Step F1.2)
- Merge
- Prepare first draft of broad outline for discussion

Readings
- Aitchison et al. D (p.17-36)

Week 7. Developing a conceptual schema and building a classified index

Project
- Group discussion (sorting into broad outline)
- Stat work on the classified index
- Develop first draft of individual sections

Readings
- Materials will be distributed

Week 8. No class (Mardi Gras)

Week 9. Group discussion: building classified index
Project

- Continue working on the classified index
- Integrating the individual sections and designing an integrated conceptual schema

Week 10. Group discussion: conceptual schema

Project
- Think about the format of thesaurus
- Continue developing the classified index
- Complete preliminary version of the arrangement within each entity type

Readings
- Thesaurus format: Soergel (1974): Sections D1-D1.6, D2 (skip D2.2), DD3-3.5

Week 11. Review discussion
- Summary of terminological and conceptual structure (Soergel B4.3)
- Concept formation in thesaurus construction
- Definitions and scope note (Soergel C3)
- Type of concepts, descriptors, terms to be included in a thesaurus (Soergel C4)
- The lead-in structure (USE and SEE) (Soergel C5)
- Synonyms proper and spelling variants (Soergel C6, C7)
- Thesaurus format and design options
- Display of descriptors and their interrelationships (Soergel D3-D3.5)
- Notation (Soergel D4)

Project
- Group discussion: classified index, thesaurus format

Readings
- Soergel (1974) sections (see above)

Week 12. Evaluation of existing thesauri (lecture + examples)
- User centered evaluation of thesaurus: retrieval and interface

Project
- Develop cross-references within and between the individual sections (in groups) and streamline the classified index

Readings
- Shiri, Ali Asghar; Revie, Crawford; Chowdhury, Gobinda. (2002) Thesaurus-enhanced search interfaces. *Journal of Information Science*. 28(2), pp.111 - 122.
- Gantzmann, J. (1990) Criteria for the evaluation of thesaurus software. *International Classification*, 17(3/4), 148-154; with appendix, p. 155-157 Check-

list for thesaurus software. (Section B parts: 3.1 'Display on the screen' and 4.2 'Retrieval')

- Owens, Leslie Ann and Cochrane, Pauline Atherton (2006). Thesaurus Evaluation, *Cataloging & Classification Quarterly*, 37: 3, 87-102

Week 13. Use of computers in thesaurus construction

- Automatic construction of thesauri
- Using thesaurus construction software (demo)
- Thesaurus software
- A thesaurus as a knowledge base. Frame-based thesaurus models.

Project
- Start working on report.
- Write introduction (Soergel 1974, Section D1.8)

Readings
- Computer assistance in thesaurus construction (Soergel 1974, Chapter G)
- Automatic methods in thesaurus construction (Soergel 1974, Chapter H)
- Riesland, Melissa A. (2006). Tools of the Trade: Vocabulary Management Software, *Cataloging & Classification Quarterly*, 37: 3, 155-176.
- Hedden, H. (2008). Comparative evaluation of thesaurus creation software. *The Indexer*, 26(2), p. 50-59.

Week 14. No class (Spring Break)

Week 15. Updating and maintenance of thesauri; Thesaurus on the Web

Project
- Do sample indexing and evaluate results

Readings
- Updating and maintenance: Soergel 1974 Chapter J.
- Compatibility and cooperation: Soergel 1974 Chapter K.
- Shiri, Ali (2000) Thesauri on the Web: current developments and trends. *Online Information Review*, Vol. 24, No 4. pp. 273 - 279.
  http://arizona.openrepository.com/arizona/bitstream/10150/105440/1/thesauri.pdf
- Hudon, Michele (2003). True and tested products: thesauri on the web. *Indexer*, Vol. 23, No. 3, pp. 115 –119.
  http://wilsontxt.hwwilson.com/pdffull/02825/axea3/6s8.pdf

Week 16. Knowledge organization systems (KOS)

- Thesauri, taxonomies, ontologies, and the semantic Web

- KOS standards
- Final review

Readings
- DESIRE information Gateway Handbook, Section 2.5 Subject Classification, Browsing and Searching. http://www.desire.org/handbook/contents.html
- Miller, Paul. Towards a typology for portals, *Ariadne* 37, October 2003. http://www.ariadne.ac.uk/issue37/miller/
- Almeida, Mauricio Barcellos and Barbosa, Ricardo Rodrigues (2009). Ontologies in Knowledge Management Support: A Case Study. *JASIS&T*, 60(10):2032–2047, 2009

Week 17. Final Exam
- Evaluation of an existing thesaurus
- Project and report due.