

LIS 419 – Indexing and thesaurus construction



Design, evaluation, and improvement of systems providing subject access to information resources. Indexing, classification and taxonomy, indexing language development, abstracting, algorithmic approaches. Subject organization and retrieval in a range of information systems and settings including Web sites and digital libraries. Practical exercises, individual or group projects, in-class presentations. 3 credits. *Prerequisites:* LIS 400, 407 and 415.

Please note that the syllabus is normally an online document, and subject to change.

Course-specific learning outcomes

By the end of the course, the student

- uses the tools of subject analysis appropriately
- uses the vocabulary of the field appropriately
- demonstrates knowledge of core concepts and principles
- identifies and evaluates the subject analysis components of an information discovery service, including user-supplied subject analysis
- determines the best approach for solving a given subject access problem
- identifies and evaluates research done in subject analysis
- creates simple subject analysis tools

Program-level student learning outcomes addressed by this course

GSLIS graduates

- communicate clearly and effectively in a range of formats to a variety of audiences
- demonstrate critical thinking in their practice of library and information science

Outline

The following is a listing of the topics covered in the course, in the order in which they are covered.

Course welcome and mechanics

- Model; Types and components of subject analysis systems, languages, and environments (broadly); Resources (including standards); Starting the project

Introduction to indexing

- What is indexing; Aboutness

Doing indexing

- The intellectual process; Using a thesaurus

A little help from the crowd

- Tagging and folksonomy

Creating subject analysis tools

- Creation, construction, and characteristics of thesauri, taxonomies, classification schemes, ontologies; Guided navigation

Evaluating indexing systems

- Types of evaluation; Performance criteria

Single-item indexing

- Creating indexes; Evaluating indexes; Indexing as a profession

Application area

- Nontextual objects (images, sounds), digital libraries

Abstracting

- Types of abstracts; Writing abstracts; Evaluating abstracts

A little help from our friends

- Automatic and machine-assisted indexing and abstracting

Wrap up

- Staying up to date; Project presentations

Readings & resources

Books & articles

There is no textbook that really covers the course perfectly, but the closest thing is: Browne, G., & Jermey, J. (2007). *The indexing companion*. Cambridge, England: Cambridge University Press. I do recommend that you purchase this one, as it is available from amazon in [paperback](#) for \$50 and [Kindle](#) for \$32. I will assign readings from this text.

The following materials are some of the classics or key pieces. I have noted where these are on reserve at Simmons. If you are a remote student, these will be available via inter-library loan, or possibly in a university library near you.

- Aitchison, J., Gilchrist, A. & Bawden, D. (2007). *Thesaurus construction and use: A practical manual* (4th ed.). London, England: Aslib. Also available as a Kindle edition for purchase or rent (ca \$25 a month) [On reserve]
- Anderson, J. D., & Pérez-Carballo, J. (2001). The nature of indexing: How humans and machines analyze messages and texts for retrieval. Part I: Research, and the nature of human indexing. *Information Processing & Management*, 37, 231-254. Retrieved from <http://www.calstatela.edu/faculty/jperezc/Publications/artAnderson2001a.pdf> [as a Simmons e-resource]; Part II: Machine indexing, and the allocation of human versus machine effort. *Information Processing & Management*, 37, 255-277. Retrieved from <http://www.calstatela.edu/faculty/jperezc/Publications/artAnderson2001b.pdf> [as a Simmons e-resource]. [On reserve]
- Baca, M. (Ed.). (2002). *Introduction to art image access: Issues, tools, standards, strategies*. Los Angeles, CA: Getty Research Institute. [On reserve]
- Booth, P. (2001). *Indexing: The manual of good practice*. München, Germany: K.G. Saur. [On reserve]
- Broughton, V. (2006). *Essential thesaurus construction*. London: Facet [On reserve]
- Chan, L. M., Richmond, P., & Svenonius, E. (Eds.). (1985). *Theory of subject analysis: A sourcebook*. Littleton, CO: Libraries Unlimited. [On reserve]
- Foskett, A. C. (1996). *The subject approach to information* (5th ed.). London, England: Library Association. [On reserve]
- Hjørland, B. (1992). The concept of subject in information science. *Journal of Documentation*, 48 (2), 172-200. Retrieved from http://db.academia.edu/BirgerHj%C3%B8rland/Papers/1343401/The_concept_of_subject_in_information_science
- Lambe, P. (2007). *Organising knowledge: Taxonomies, knowledge and organisational effectiveness*. Oxford, England: Chandos. [This is from the knowledge management point of view, so if that's your career plan, this would supplement Aitchison & Gilchrist on thesaurus construction]
- Lancaster, F.W. (2003). *Indexing and abstracting in theory and practice* (3d ed.). Champaign, IL: University of Illinois, Graduate School of Library & Information Science. [This used to be the text for the course but it is sadly out of print now] [On reserve]

- Latthrop, L. (2000). Index usability test questions. *A-Z: The Indexing SIG Newsletter*. Retrieved from <http://www.stcsig.org/idx/articles/usability.pdf>
- Mulvaney, N. (2005). *Indexing books* (2d ed.). Chicago, IL: University of Chicago Press. Also a Kindle book for under \$10. [One of the best known indexing manuals] [On reserve]
- Rafferty, P., & Hilderly, R. (2004). *Indexing multimedia and creative works: The problems of meaning and interpretation*. Burlington, VT: Ashgate.
- Roe, S. K., & Thomas, A. R. (Eds.). (2004). *The thesaurus: Review, renaissance and revision*. New York: Haworth. [A special issue of *Cataloging & Classification Quarterly*, 37(3/4). Retrieved from <http://0-www.tandfonline.com.library.simmons.edu/toc/wccq20/37/3-4> [as a Simmons e-resource]
- Saller, C. (2012, May 1, 9, and 14). Book indexing [A three-part entry in the Chronicle of Higher Education's Lingua Franca blog]. Part 1 retrieved from <http://chronicle.com/blogs/linguafranca/2012/05/01/book-indexing-part-1-is-a-computer-the-right-person-for-the-job/>. Part 2 retrieved from <http://chronicle.com/blogs/linguafranca/2012/05/09/book-indexing-part-2-infinite-loops-and-easter-eggs/>. Part 3 retrieved from <http://chronicle.com/blogs/linguafranca/2012/05/14/book-indexing-part-3-tips-for-do-it-yourselfers/>
- Salo, D. (2009, August 5). The humble index [Web log post]. Retrieved from <http://scienceblogs.com/bookoftrogool/2009/08/25/the-humble-index/>.
- Schwartz, C. (2008). *Thesauri and facets and tags, oh my! A look at three decades in subject analysis*. *Library Trends*, 56, 830-842. Retrieved from http://0-muse.jhu.edu.library.simmons.edu/journals/library_trends/v056/56.4.schwartz.html [as a Simmons e-resource]
- Smith, S.L., & Kells, K. (2005). *Inside indexing: The decision-making process*. Bend, OR: Northwest Indexing Press. [Great insights into the actual processes involved in book indexing] [On reserve]
- Spiteri, L. (1998). A simplified model for facet analysis. *Canadian Journal of Information and Library Science*, 23(Apr-Jul), 1-30. Reprinted by the Information Architecture Institute. Retrieved from http://iainstitute.org/en/learn/research/a_simplified_model_for_facet_analysis.php
- *Statistical association methods for mechanized documentation: Symposium proceedings*. (1965). Washington, DC: GPO. [This is the conference where IR began]
- Taylor, A. G. & Joudrey, D. N. (2009). *The organization of information* (3d ed.). Westport, CT: Libraries Unlimited. [Course textbook for LIS 415]
- Tudhope, D., Koch, T., & Heery, R. (2006). *Terminology services and technology: JISC state of the art review*. Retrieved from <http://www.ukoln.ac.uk/terminology/TSreview-jisc-final-Sept.html> [A terrific summary of the varieties of terminologies out there]
- Tunkelang, D. (2009). *Faceted search*. Morgan & Claypool. (Synthesis Lectures on Information Concepts, Retrieval, and Services). [On reserve]

- Van Rijsbergen, C. J. (1979) *Information retrieval* (2d ed.). London, England: Butterworths. Retrieved from <http://www.dcs.gla.ac.uk/Keith/Preface.html>
- Wichowski, A. (2009). Survival of the fittest tag: Folksonomies, findability, and the evolution of information organization. *First Monday*, 14(5). Retrieved from <http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/2447/2175>
- Wilson, P. (1968). *Two kinds of power: An essay on bibliographical control*. Berkeley, CA: University of California Press. (Librarianship, 5). See especially Chapter V. Subjects and the sense of position (pp. 69-92). Retrieved from http://books.google.com/books?id=DePy_aazKl4C&pg=PA69&source=gbs_toc_r&ad=4#v=onepage&q&f=false[This is a Google book, and I don't know whether they will snatch access away at any point, so let me know if they do.]
- Zafran, E. L. (2005). *Index it right! Advice from the experts*. New York: American Society of Indexers. [Sections on indexing philosophy, biography, theology, art, horticulture, encyclopedias, computer manuals, and Web sites] [On reserve]

Standards

Throughout the course you should examine the latest version of whatever standards pertain to the topic under discussion and to your project.

- Anderson, J. D. (1997). *Guidelines for indexes and related information retrieval devices*. Bethesda, MD: NISO Press. (NISO technical report TR02-1997). Retrieved from <http://www.niso.org/publications/tr/tr02.pdf>
- International Organization for Standardization. (2004). *Abstracts for publications and documentation. ISO 214: 1976*. Geneva, Switzerland: International Organization for Standardization.
- International Organization for Standardization (ISO). (1996). *Guidelines for the content, organization and presentation of indexes. ISO 999: 1996*. Geneva, Switzerland: International Organization for Standardization. [Also BS 999: 1996]. [On reserve]
- International Organization for Standardization. (1985). *Methods for examining documents, determining their subjects, and selecting index terms. ISO 5963-1985(E)*. (Also BS 6529: 1984) [On reserve]
- International Organization for Standardization. (2011). *Thesauri and interoperability with other vocabularies. Part 1: Thesauri for information retrieval. ISO 25964-1:2011*. Geneva, Switzerland: International Organization for Standardization. [Replaces 2788: 1986 and 5964: 1985; Also BS 25964-1:2011] [On reserve]
- National Information Standards Organization. (1997). *Guidelines for abstracts. ANSI/NISO Z39.14-1997 (R2009)*. Bethesda, MD: NISO Press. Retrieved from http://www.niso.org/apps/group_public/project/details.php?project_id=32
- National Information Standards Organization. (2005). *Guidelines for the construction, format, and management of monolingual thesauri. Z39.19-2005 (R2010)*. Bethesda, MD: NISO Press. Retrieved from http://www.niso.org/apps/group_public/project/details.php?project_id=46[Also ISO 2788-1986(E)]

Unusual print thesauri & other tools

Web-accessible thesauri are listed below at Subject analysis: Indexing languages.

- Boggs, B. B., & Patterson, D. W. (1984). *An index to selected folk recordings*. Chapel Hill, NC: University of North Carolina Press.
- Burger, H. G. (1984). *The wordtree*. Merriam, KS: The Wordtree, Inc.
- Chenhall, R. G. (1988). *The revised nomenclature for museum indexing*. Nashville, TN: AASLH Press.
- Garnier, F. (1984). *Thésaurus iconographique*. Paris, France: Le Léopard d'Or.
- Petersen, T., & Barnett, P. J. (1994). *Guide to indexing and cataloging with the Art and architecture thesaurus*. New York: Oxford University Press.
- Harrold, A., & Lea, G. (1991). *Musaurus: A music thesaurus*. London, England: Music Press.
- Knapp, S. D. (2000). *The contemporary thesaurus of search terms and synonyms: A guide for natural language computer searching*. Phoenix, AZ: Onyx.
- Olderr, S. (1991). *Olderr's fiction subject headings: A supplement and guide to the LC thesaurus*. Chicago, IL: American Library Association. [On reserve]
- Royal Commission on the Historical Monuments of England. (1995). *Thesaurus of monument terms*. London, England: The Commission.
- Unesco (1995). *Unesco thesaurus*. Paris, France: Unesco.
- Waal, H. van de. (1985). *Iconclass*. Amsterdam, Netherlands: North-Holland, 1985. [The scheme is in the index volumes, and there is a user manual: Straten, Roelof von. (1994). *Iconography, indexing, Iconclass: A handbook*. Leiden, The Netherlands: Folcor.]

Journals

I keep a list of [journals](#) in this and related fields in Pinboard. Note that many of the print journals are [available in full text](#) through one or another of our [e-resources](#).

Web resources

The following list is on my [LIS resources page](#), which also includes links to other relevant resources for doing research in LIS.

- [key agencies and scholars](#) in information retrieval
- [indexing resources](#) and [indexing languages](#)
- [classification resources](#) and [applications of classification](#)
- applications of [guided navigation](#), [visualization](#), and [folksonomy](#)
- [linked data and semantic web](#)
- related topics: [digital identity](#)

Student FAQ

- How do I get in touch with you?
- The best thing to do is to e-mail me at candy.schwartz@simmons.edu. Please use your Simmons e-mail – I get too much spam to spot your message if it is not from your simmons.edu address. My office phone number is 617-521-2849, and you can leave voice mail there if I am not in. My office is P-310C – turn left when you come through the glass doors to P-310. I try not to work at home, so do not rely on me to be available between Friday afternoon and Monday morning, though I usually am. I can frequently be found via Twitter (candyschwartz), Facebook, AIM (candysschwartz), Google chat (candy.schwartz), and Skype (candyschwartz).
- When may I see you face to face?
 - Anytime you can find me. Usually between 7 AM and 5 PM on weekdays (sometimes I work at home on Mondays). Don't worry about whether it's my "office hours" or not – if my door is open, just drop in. For online classes, I will hold virtual office hours, but am usually accessible at other times.
- May I have a copy of the course outline?
 - See appropriate pages in Candy@
- I don't have the prerequisites (or I am taking a prerequisite course simultaneously) – will you let me in?
 - You must have fulfilled prerequisites **prior to the first class meeting**, or you must have "permission of instructor". In all other cases I reserve the right to ask you to drop the class.
- I am wait-listed – will you let me in?
 - No. But someone may withdraw after the first class, so do please come to the first class in case you make it in.
- May I audit the class unofficially?
 - Sorry, no. You may not attend classes unless you are registered for the course. We do have an official "audit", for alumni/ae, for less than the normal cost of the course and for no credit. Students about to graduate may also register for an "extra course" at a reduced cost.
- Is your syllabus like a contract between you and me?
 - No. The syllabus is not a contract. I reserve the right to alter the course requirements and/or assignments based on new materials, class discussions, or to meet specific pedagogical objectives.
- Will there be a lot of computer use in the course?
 - I expect new and returning students to have mastered the concepts presented in the [Technology Orientation](#) and in any course prerequisites.
- Can I bring a computer/netbook/ipad/smartphone to class?
 - I expect phones of all kinds to be turned off for calls unless you have small children. I expect you to be courteous in using computers – take notes and follow links mentioned in class, but don't do e-mail, social networking, games, etc. If the class bores you that much, then don't come. If you choose

to ignore this request, then do it quietly. Respect me and respect the people around you who are trying to learn. I will call you out in class if you snicker over a screen when I am not being funny (and I do try to be funny from time to time), and especially if you are sharing with your neighbour.

- What is eLearning, or, Does this course use eLearning?
 - All of my courses have an e-learning component of one kind or another. Simmons uses the [Moodle](#) learning management system. Use the same username and password as you do for logging in to Webmail.
 - On the use of dropboxes: If you mess up and you need to repost an assignment, you will need to e-mail me and I will need to take several steps. So please make sure you have done all of your proofreading and spellchecking before you submit. **Also, submit all of your files at once if you have a multi-file assignment.** When I return assignments, I use the same dropbox. Often I make comments. They may include “and if you do x, then I will give you a better grade”. So don’t just look at your grade – look at the comments too.
- Do you have preferences for file names and formats?
 - Please submit any documents which might be commented on by me in either .doc or .rtf – it makes it easier for me to comment and correct. Sometimes I will specifically request an assignment in another format.
 - For any assignment file you turn in, whether through an eLearning system or by e-mail: When you name your files, could you please follow the pattern of (all lowercase): asst[number]-[surname].doc (or .rtf, .pdf., .odt, .docx, etc.). For example, if I were turning in Assignment 3, it would be asst3-schwartz.doc. If your assignment consists of more than one file, make a letter for the last element before the extension (e.g., asst3-schwartz-a.doc, asst3-schwartz-b.doc, and so on). Don’t put in spaces, don’t put in periods other than before the extension, don’t use brackets or underbars, and make sure that the extension is present. (BTW: Using spaces or extraneous dots in filenames is a bad habit all around, not just for me, but for many other cases where you are sharing files.) By default, Windows “shields” you from seeing file extensions. If you are a Windows-using information professional, you should set folder options so that you see the entire filename. To change the default, choose Start -> Control Panel -> Folder Options -> View, and then **uncheck** “Hide extensions for known file types”.
 - If you use Open Office and your document has tables, there’s a good possibility that they will be trashed by Word (which I use). If you create a file with screen captures using a Mac-based word processor and you save as Word, there’s a good possibility that I will not be able to see the pictures. The easiest way to check is to open these files using Word on some machine other than yours.
- How much time will the course take outside of class?
 - Lots. I usually estimate 10-15 hours outside of class per week, although it varies from course to course, semester to semester, and week to week.
- What books should I buy?

- Don't buy books until after the first class. Try to find books second-hand from students who have taken the course. Consider [amazon.com](https://www.amazon.com), [Barnes & Noble](https://www.barnesandnoble.com), or interlibrary loan. Texts are listed in the course Web pages and are on reserve, usually in multiple copies.
- I have to miss a class – what should I do?
 - First class: Leave me voice mail or e-mail (see [above](#)). As soon as possible, get first day materials. It would be nice if you could get someone's notes.
 - Not the first class: Attendance is only required for the first two class meetings. After that, if you cannot be present, do not bother to tell me. Since assignments are turned in electronically, your absence should not affect your ability to turn an assignment in on time. Call a fellow student and arrange to get notes from the day. You will miss any points based on in-class work.
- May I tape your lectures?
 - Sure, but I won't wear a microphone.
- What is your grading system?
 - A: Superior performance in all aspects of the work; exhibits mastery of the subject matter as well as qualities such as analytical ability, originality, creativity, and interpretation; a great job, well presented.
 - A-: Extra effort was apparent; excellent performance; superior achievement in most aspects of the work.
 - B+: Strong, solid performance; above basic course requirements.
 - B: Good work; generally meets basic requirements, but with some problems.
 - B- and lower: Does not attain acceptable levels in significant aspects of the assignment/course; sins of omission or commission.
 - A/A-, A-/B+, B+/B, etc.: Borderline grades
 - Each grade has a point value (e.g., A is 4 and A- is 3.67). I keep grades in an Excel spreadsheet and calculate an average. The final course grade takes into account this average and other course-specific criteria.
- May I turn an assignment in late?
 - Generally yes (and sometimes no), but not without penalties. Late assignments will receive a lower grade than they would have otherwise, and the later they are handed in the lower the grade. I set a deadline in the elearning system, and after that things are noted as being late. You do not need to tell me that an assignment is going to be late (I may ask where it is when I grade), nor do you need to explain why, just get it in as soon as possible. When someone asks if they can hand something in late, I usually say okay. That means that I will accept it, but it still counts as late.
- What happens if I don't do all the assignments?
 - All assignments and projects are required for completion of a course. At some point I will start sending you notes so that you and I both know that you have been warned about what is missing. I will keep a record of these notes.
- What style manual and format do you prefer for assignments?

- References and in-text citations: I prefer the APA style, but as long as you use a single style manual consistently, I don't care which. I have bookmarked some resources in Pinboard on [writing and style](#).
- Essays should have titles, headings, subheadings, introductions, conclusions, and other devices reflecting organized thought, and I prefer single-spaced layout. Careful proofreading for both grammar and spelling is suggested – presentation counts. Learn to capture screens and incorporate them into papers.
- I know that this practice is common for other faculty, but I prefer that you not include the questions when you turn in an assignment. I know what the questions are – I just need to see the question numbers and the answers.
- Here are my [pet peeves](#) as far as writing is concerned. I don't care if common practice has made some of these acceptable – they are not acceptable to me. It's a short list, but growing.
- On the subject of cutting and pasting – I can tell when your words are not yours, but something from some article you found online. And it's easy to go to a search engine or aggregator and search for the words and phrases and find the original. So if you use other people's words, put them in quotes and attribute them (see the [honour code info](#) below) – doing otherwise can be grounds for dismissal from GSLIS. Ditto (without the quotes) if you use other people's ideas.
- Do not put hard copy assignments in binders, folders, plastic sleeves, envelopes, etc.
- A group paper or project is treated as one coherent whole, not individual contributions.
- May I e-mail assignments to you?
 - You may attach assignments to e-mail **only** if I specifically ask for them that way, or if you can't turn them in in the requested manner.
- How do I get my papers back?
 - E-assignments will be returned in some e-way, usually in an elearning dropbox. Anything else goes back to your student folder in the lounge. Make sure you have a folder.
- Will you keep copies of my assignments?
 - Normally I keep nothing without asking. However, unless you indicate otherwise by e-mail to me, I do reserve the right to keep copies of your work for use in accreditation activities. For example, copies of student work will be collected for the Massachusetts Department of Elementary and Secondary Education accreditation process. Your name will be removed from the work prior to use.
- May I have an extension to complete the course?
 - My grades are usually due within a day or two after the last class. There is no such thing as an extension. Incompletes are only granted for medical emergencies and disasters of a similar nature. This is school policy, and I may not violate it. You may withdraw from a class up **until** the last meeting.

- Will you write me a letter of reference?
 - After the class is over, if you would like me to act as a reference for you, you will first need to set up a file in the [Career Education Center](#), and I will then deposit your reference in that file, so it can be sent to prospective employers at your request. Then I will need the following: an authorization form from the Center (don't forget to sign the form), your up to date résumé, a list of the courses you took at Simmons (and **which ones in what semesters were with me**), and a brief statement of career objectives. I will only do telephone references if you have done this.
- Does Simmons have an honour code?
 - Simmons College expects each of its students to uphold the [Honor Code](#). Major violations are cheating and plagiarism. Cheating includes copying others' works, collaborating without authorization, and accessing others' computer files without authorization. Plagiarism includes intentionally or unintentionally using someone else's words, works, thoughts, or expression of ideas without giving proper credit (also see the cutting and pasting note under the style manual discussion above). Unauthorized copying of print or non-print materials (including machine-readable data) is a violation of the laws governing intellectual property. Deliberately altering configuration files, batch files, and other installed software on computers in the Tech Lab is unprofessional and inconsiderate.
- I have a physical or cognitive challenge – what should I do?
 - Reasonable accommodations will be provided for students with documented physical, sensory, systemic, cognitive, learning and psychiatric disabilities. If you have a disability and anticipate that you will need a reasonable accommodation in this class, it is important that you contact the [Center for Academic Achievement](#) (617-521-2474) early in the semester so that accommodations can be provided. Students with disabilities are also encouraged to contact their instructors to discuss their individual needs for accommodations. Faculty cannot provide accommodation without correspondence from the Center. The Center maintains strict confidentiality. The Center also has writing tutors.