

Information Studies 293, UCLA
Fall, 2006
Thursdays, 9:00am-12:30pm
GSE&IS Room 245

Information Retrieval PhD Seminar

Christine L. Borgman, Professor & Presidential Chair in Information Studies
Department of Information Studies
235 GSE&IS Bldg.
310/825-6164
borgman@gseis.ucla.edu

This course is offered in alternate years as one of the six seminars for the Information Studies PhD program. It was last taught in Winter, 2005 by Jonathan Furner. The syllabus draws heavily on Prof. Furner's course design.

Catalog description

“Seminar, four hours. Designed for Ph.D. students. Intellectual principles for organization of information, including principles for design of systems for acquiring, organizing, and retrieving information. Also includes system-specific user studies to the extent that design of information systems is predicated on their evaluation and use. S/U or letter grading.”

Objectives

The aim of this course is to provide an introduction to some of the principal research issues in the doctoral specialization area of information retrieval (a.k.a. information systems). It is intended to lay the foundation for doctoral students' preparation in this area before proceeding to the Written Qualifying Examination.

Outcomes

Upon satisfactory completion of this course, you will have demonstrated:

- your knowledge and appreciation of the history of research in the field, of its philosophical foundations, of its core literature, and of participating scholars and institutions;
- your ability to identify, analyze, and articulate significant research problems;
- your ability to apply appropriate research methodologies to the problems identified; and
- your understanding of key theoretical and practical contributions to the field.

Course audience, prerequisites

The course is intended for PhD students in Information Studies. Others may enroll with instructor's permission if they have background commensurate with IS PhD students. This course addresses history and theory, not practice. Those interested in practical aspects of the design of information retrieval systems should take the master's level courses in information retrieval, information technology, digital libraries, digital archives, and systems analysis and design.

Grading procedures and policy

See separate document for details of assignments

40% In-class presentations and short papers (2 @ 20%)
40% Seminar paper
20% Class participation

Readings are to be completed prior to the class meeting for the week. The grade for class participation will be based on preparation for the discussion, contributions to the discussion, and attendance. The course is conducted as a research seminar.

Late assignments will be marked down 2 points per day late. **No assignments will be accepted after 5pm Thursday, December 14 without advance permission from the instructor.**

Course materials

Course materials are in three categories (bibliographic references at end of document)

1. Required for purchase:

Sparck Jones, K. and P. Willett, Eds. (1997). Readings in Information Retrieval. San Francisco, Morgan Kaufmann. Available at LuValle Commons Bookstore.

2. Available online or in the UCLA libraries:

Readings not in the text are available online, either through the UCLA libraries or free on the World Wide Web. Those only in print are being scanned and made available in the online reserve system. To access the UCLA library databases from off campus, you must have a proxy server or VPN client set up. Most of the older articles also are available in print in the UCLA libraries.

Note that printing resources in the MIT lab are limited. Please print at home or read online (an inexpensive laser printer is a good investment for graduate school).

3. Other Resources

A. Materials from Prof. Furner's master's-level class in information retrieval (INFS 276):

- "Resources" document listing texts, journals, conferences, information services, and systems, with which you should be familiar: <http://polaris.gseis.ucla.edu/jfurner/02-03/276/276res.doc>
- Reading list: <http://polaris.gseis.ucla.edu/jfurner/02-03/276/276syl.doc>
- PowerPoint slides and handouts on individual topics:
<http://polaris.gseis.ucla.edu/jfurner/02-03/276/276.html>

B. Many texts on information retrieval offer good background reading; they are useful additions to your personal reference library (Belew, 2000; Ellis, 1996; Ingwersen & Jarvelin, 2005; Korfhage, 1997; Lesk, 2005; Meadow, 1992; Meadows, 2001; Tait, 2005; Voorhees & Harman, 2005). Yet more texts are listed in the Sparck Jones and Willett book and in Prof. Furner's resources document.

Office hours

Tuesdays, 2:30-4:30PM, GSE&IS 235. Please sign up in advance, as office hours may be full. Additional hours by appointment. Short questions can be answered by email, but discussions of term paper topics are best for office hours.

Topical course outline

Week 0 (Sept 28):	Introduction; History
Week 1 (Oct 5):	History and background
Week 2 (Oct 12):	Key concepts
Week 3 (Oct 19):	Evaluation
Week 4 (Oct 26):	Models
Week 5 (Nov 2):	No class; CB at conferences (work on term papers)
Week 6 (Nov 9):	Techniques, part 1
Week 7 (Nov 16):	Techniques, part 2
Week 8 (Nov 23):	Thanksgiving (work on term papers)
Week 9 (Nov 30):	Systems

Week 10 (Dec 7): Relevance revisited

Exam week (Dec 14): Cognitive paradigm; user studies

***Hints for finding articles below:**

1. If the <doi> links do not open automatically, go to <<http://dx.doi.org>> and enter the entire DOI citation in the text box provided.
2. The Wiley Interscience articles requires the use of cookies. You may get a "Session Cookie Error" when you first open the links. Simply close the "Error" window and open the link again. The article should appear the second time you open it.

Week 0 (Sept 28): Introduction and History

(Sparck Jones & Willett, 1997):

Preface

Chapter 1: Overall introduction

Chapter 2: History – Joyce & Needham; Luhn; Doyle

Bush, V. (1945). "As We May Think." *Atlantic Monthly*: 101-108.

URL: <http://www.georgetown.edu/faculty/jod/texts/vannevar.bush.html>

(original is available at www.theatlantic.com; UCLA proxy or VPN required)

Week 1 (Oct 5): History

(Sparck Jones & Willett, 1997):

Chapter 2: History – Maron & Kuhns; Cleverdon; Salton & Lesk

Chapter 9: Envoi - Swanson

Ellis, D. (1984). Theory and explanation in information retrieval research. *Journal of Information Science*, 8(1), 25-38.

URL forthcoming.

Presentation topics to be assigned – identify your preferences in advance of class.

Week 2 (Oct 12): Key concepts

(Sparck Jones & Willett, 1997):

Chapter 3: Key concepts

Week 3 (Oct 19): Evaluation

(Sparck Jones & Willett, 1997):

Chapter 4: Evaluation

Term paper proposals due.

Week 4 (Oct 26): Models

(Sparck Jones & Willett, 1997):
Chapter 5: Models

Student presentations

Week 5 (Nov 2): No class; CB at conferences – work on your term papers

Week 6 (Nov 9): Techniques, part 1

(Sparck Jones & Willett, 1997):
Chapter 6: Techniques – Porter; Strzalkowski; Salton & Buckley; Sparck Jones; Croft & Harper

Student presentations

Week 7 (Nov 16): Techniques, part 2

(Sparck Jones & Willett, 1997):
Chapter 6: Techniques – Robertson & Walker; Salton & Buckley; Griffiths, Luckhurst, & Willett

Sparck Jones, K., Walker, S., & Robertson, S. E. (2000a). A probabilistic model of information retrieval: Development and comparative experiments, part 1. *Information Processing & Management*, 36(6), 779-808.

URL: [doi:10.1016/S0306-4573\(00\)00015-7](https://doi.org/10.1016/S0306-4573(00)00015-7)

Sparck Jones, K., Walker, S., & Robertson, S. E. (2000b). A probabilistic model of information retrieval: Development and comparative experiments, part 2. *Information Processing & Management*, 36(6), 809-840.

URL: [doi:10.1016/S0306-4573\(00\)00016-9](https://doi.org/10.1016/S0306-4573(00)00016-9)

Student presentations

Week 8 (Nov 23): Thanksgiving

Week 9 (Nov 30): Systems

(Sparck Jones & Willett, 1997):
Chapter 7: Systems

Student presentations

Week 10 (Dec 7): Relevance revisited

Cuadra, C. A., & Katter, R. V. (1967). Opening the black box of 'relevance'. *Journal of Documentation*, 23(4), 291-303.
URL forthcoming.

Schamber, L., Eisenberg, M. B., & Nilan, M. S. (1990). A re-examination of relevance: Toward a dynamic, situational definition. *Information Processing & Management*, 26(6), 755-776.
URL: [doi:10.1016/0306-4573\(90\)90050-C](https://doi.org/10.1016/0306-4573(90)90050-C)

Harter, S. P. (1992). Psychological relevance and information science. *Journal of the American Society for Information Science*, 43(9), 602-615.
URL: <http://www3.interscience.wiley.com/cgi-bin/abstract/10049714/ABSTRACT>

Harter, S. P. (1996). Variations in relevance assessments and the measurement of retrieval effectiveness. *Journal of the American Society for Information Science*, 47(1), 37-49.
URL: <http://www3.interscience.wiley.com/cgi-bin/abstract/57639/ABSTRACT>

Barry, C. L., & Chamber, L. (1998). Users' criteria for relevance evaluation: A cross-situational comparison. *Information Processing & Management*, 34(2/3), 219-236.
URL: [doi:10.1016/S0306-4573\(97\)00078-2](https://doi.org/10.1016/S0306-4573(97)00078-2)

Student presentations

Exam week (Dec 14): Cognitive paradigm; user studies; course wrapup

(Sparck Jones & Willett, 1997):
Chapter 5: Models -- Belkin, Oddy, & Brooks

Borgman, C. L. (1986). The user's mental model of an information retrieval system: An experiment on a prototype online catalog. *International Journal of Man-Machine Studies*, 24(1), 47-64.

(Reprinted as Borgman, C. L. (1999). The user's mental model of an information retrieval system: An experiment on a prototype online catalog. *International Journal of Human-Computer Studies*, 51, 435-452.)
URL: [doi:10.1006/ijhc.1985.0318](https://doi.org/10.1006/ijhc.1985.0318)

Bates, M. J. (1990). Where should the person stop and the information search interface start? *Information Processing & Management*, 26(5), 575-591.
URL: [doi:10.1016/0306-4573\(90\)90103-9](https://doi.org/10.1016/0306-4573(90)90103-9)

Marchionini, G. (1992). Interfaces for end-user information seeking. *Journal of the American Society for Information Science*, 43(2), 156-163.
URL: <http://www3.interscience.wiley.com/cgi-bin/abstract/10049647/ABSTRACT>

Other readings: (Crane, 2006; Lynch, 2006; White & McCain, 1998)

White & McCain: <http://www3.interscience.wiley.com/cgi-bin/fulltext/28049/PDFSTART>
Lynch: www.cni.org

REFERENCES

- Belew, R. K. (2000). Finding out about: A cognitive perspective on search engine technology and the WWW. Cambridge, UK: Cambridge University Press.
- Crane, G. (2006). What do you do with a million books? *D-Lib Magazine*, **12**(3). Retrieved from <http://www.dlib.org/dlib/march06/crane/03crane.html> on 17 August 2006.
- Ellis, D. (1996). Progress and problems in information retrieval (2nd ed.). London: Library Association.
- Ingwersen, P. & Jarvelin, K. (2005). The Turn: Integration of Information Seeking and Retrieval in Context. Dordrecht, Netherlands: Springer.
- Korfhage, R. R. (1997). Information storage and retrieval New York: Wiley.
- Lesk, M. (2005). Understanding Digital Libraries (2nd ed.). San Francisco: Morgan Kaufman.
- Lynch, C. A. (2006). Open computation: Beyond human-reader-centric views of scholarly literatures. *Open Access: Key strategic, technical and economic aspects*. Jacobs, N. Oxford, U.K., Chandos.
- Meadow, C. T. (1992). Text Information retrieval systems. San Diego: Academic Press.
- Meadows, J. (2001). Understanding Information. Munchen: K. G. Saur.
- Sparck Jones, K. & Willett, P. (Eds.). (1997). Readings in Information Retrieval. San Francisco: Morgan Kaufmann.
- Tait, J. I. (Ed.). (2005). Charting a new course: Natural Language Processing and Information Retrieval, Essays in Honour of Karen Sparck Jones. Dordrecht, Netherlands: Springer.
- Voorhees, E. M. & Harman, D. K. (Eds.). (2005). TREC: Experiment and Evaluation in Information Retrieval. Cambridge, MA: MIT Press.
- White, H. D. & McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972-1995. *Journal of the American Society for Information Science*, **49**(4): 327-355.