

LBSC 708X/INFM 718X: Seminar on E-Discovery

Spring 2009, Mondays, 6:00-8:45 PM

3 Credit Hours

Short Description:

Information retrieval concepts in the applied context of litigation, investigations and review of electronically stored information for public release. Discussion of legal requirements, professional standards, best practices, technical capabilities for automated support, evaluation, and risk management. Intended for students interested in legal, corporate or institutional settings involving provision of responsive controlled access to large collections of electronic records.

Faculty:

Jason R. Baron, Esq., iSchool (Adjunct Professor), National Archives and Records Administration (Director of Litigation), The Sedona Conference® Working Group on Electronic Records (Steering Committee); Text Retrieval Conference (Legal Track Coordinator), jason.baron@nara.gov

Douglas W. Oard, Ph.D., iSchool (Associate Dean for Research), University of Maryland Institute for Advance Computer Studies, Text Retrieval Conference (Legal Track Coordinator), oard@umd.edu

Students:

The course is designed for students in the MLS program with interests in archives, records management, e-government, information access, information technology, for students in the MIM program with interests in the intersection between information management and legal and regulatory requirements, and Ph.D. students who are interested in the substantial number of open research questions in this field.

Prerequisites:

Those standard for any 700-level course (briefly: for MLS and MIM students, completion of, or simultaneous registration for, all core courses; or permission of the instructor).

Relationship with other courses:

This course interleaves with the content of LBSC 682 (Management of Electronic Records). Essentially, LBSC 708X/INFM 718X picks up where that course leaves off. Our starting point is that there exist some records, perhaps well-managed, perhaps not, and we now need to provide access to those records. There is considerable synergy with LBSC 781 (Access Techniques and Systems for Archives). We focus on evidentiary uses of archival (and other) materials in a legal context, (mostly) on born-digital materials, and on making do as best we can with the records in the state that we find them. LBSC 781 includes a broader range of materials, users, and uses (notably, scholarly access), and a focus on the arrangement and description issues that are important for best supporting future access. There is little content overlap at all with LBSC 764 (Access to Legal Information), although students with an interest in the law might find both courses to be of interest. LBSC 764 focuses on providing access to legislation, regulation, and case

law. By contrast, we focus on access to digital evidence as required in various legal and regulatory contexts.

Background:

Civil litigation increasingly involves requests on the part of one or both opposing parties in a lawsuit for responsive documents from vast corporate or institutional storehouses of electronically stored information (ESI). I-school graduates in institutional settings (e.g., corporate or government) may well be asked to take a leading role in providing access to ESI for the purpose of responding to legal and investigatory demands, or (in government) in response to Freedom of Information Act (FOIA) requests. Searching enormous collections of sensitive information (e.g., email, word processing, dynamic databases, instant messaging, or intranet wikis) from a combination of operational information systems (e.g., networks, personal hard drives, and PDAs), records management systems, and materials produced originally for disaster recovery rather than to support future access poses challenges that the legal informatics profession is just beginning to grapple with.

Goals:

- Understand the legal frameworks by which access to sensitive information is managed in civil discovery, regulatory and legislative investigations, and FOIA requests in the USA and in comparable settings in some other parts of the world.
- Develop an analytical framework for e-discovery and related problems that integrates organizational, technical, legal, and procedural perspectives.
- Explore current issues in e-discovery.

Tentative List of Topics:

- The role of electronic evidence in American society
- Evidence and RM: The digital document life cycle in institutional settings
- Litigation: civil discovery; regulatory and congressional investigations
- Technical framework: support for search and sensemaking
- Case study: Email e-discovery in US vs. Philip Morris et al.
- Email archiving and other forms of electronic evidence management systems
- Special topic: Digital management of born-on-paper records
- Special topics: E-FOIA and the Armstrong case
- Special topics: Databases, transaction records, and usage logs
- Special topics: Computer forensics and communications law
- Risk management
- Structure of the professions (records management, evidence management, law)
- International perspectives (EU, Japan, China)

Learning Methods:

- A typical class session will have the following structure:
 - Introductory presentation (~45 minutes)
 - Response by a designated respondent (~15 minutes)
 - Discussion based on presentation (~15 minutes)
 - In-class exercise (~45 minutes)
 - Synthesis discussion (~30 minutes)
- The introductory presentation will normally be given by one of the instructors or by a guest speaker, as appropriate to the topic. We will adjust the schedule to accommodate the availability of guest speakers with (for example) experience as:
 - Judge and/or magistrate
 - Legal practitioner
 - Government official
 - Expert on information retrieval
- The response will be presented by an instructor during the first few weeks, and by a student in subsequent weeks.
- Initial in-class exercises will be instructor-designed and instructor-led, usually in small groups. Some in-class exercises later in the semester will be student-designed and student-led (with advance planning support from an instructor). Students will normally work in teams of 2 or 3 for this task.
- Weekly preparation will include at least one assigned reading related to the introductory presentation and a requirement to find and read background materials on at least on real-world event or issue related to that week's topic (e.g., for week 3, some students might choose to read about the release of 3,000 Justice department emails for the congressional investigation of the US Attorney firings). Some topics will be suggested for each week (which students may go beyond), but each student will be expected to identify their own sources.
- Four homework exercises are planned:
 - Inventory your personal electronic "records"
 - Complete a modified version of the TREC legal track interactive task
 - Select an electronic evidence management system for a hypothetical case
 - Prepare an actual E-FOIA request for records of personal interest to you
- A term paper on some aspect of e-discovery will be required.

Grading:

The grading strategy has been selected to maximize the incentive for individual engagement with aspects of the topic that are of greatest interest to each student.

- Homework: 40% (10% each)
- In-class leadership (as discussant or as in-class exercise leader): 10%
- Other contributions (in class, on the email list, as a Blog or Wiki): 10%
- Term paper: 40%

Readings:

The reading list is not yet finalized, but this list is representative of the types of readings that we expect students to select from. Case law, professional practice and technical capabilities are evolving rapidly, so this list will be updated prior to the start of the semester and as we go along.

Armstrong v. Executive Office of the President, 1 F.3d 1274 (D.C. Cir. 1993)

Axelrod, Robert, *The Evolution of Cooperation*

Bailey, P., de Vries, A.P., Craswell, N., Soboroff, I., "Overview of the TREC 2007 Enterprise Track," *The Sixteenth Text Retrieval Conference (TREC 2007) Proceedings*, Gaithersburg, MD, November, 2007. <http://trec.nist.gov>.

Baron, J., "The TREC Legal Track: Origins and Reflections on the First Year," *Sedona Conference Journal*, vol. 8, pp. 251-259, 2007

Baron, J., "Information Inflation: Can the Legal System Adapt?", 13 RIC. J.L. & TECH. 10 (2007), <http://law.richmond.edu/jolt/v13i2/article10.pdf>. (with George L. Paul, co-author)

Baron, J., "Toward a Federal Benchmarking Standard for Evaluating Information Retrieval Products Used in E-Discovery," *Sedona Conference Journal*, vol. 6, pp. 237-246, 2005 (available from Westlaw and LEXIS)

Baron, J., Lewis, D., and Oard, D. "TREC 2006 Overview"; <http://trec-legal.umiacs.umd.edu> (and other selected readings from TREC 2006/2007)

Baron, J., "The PROFS Decade: NARA, Email and the Courts," in *Thirty Years of Electronic Records* (Bruce Ambacher, ed. 2003)

Baron, J. "Email Metadata in a Post-Armstrong World," Presentation at 3rd IEEE Metadata Conference, available at www.archives.gov/era/pdf/baron-email-metadata.pdf

Blair, D., and Maron, M. "An Evaluation of Retrieval Effectiveness for a Full-Text Document-Retrieval System," *Communications of the ACM*, 28(3), pp. 289-299, 1985.

Blair, D., "Wittgenstein, Language, and Information: 'Back to the Rough Ground!'" (2006)

Boehning, H.C., and Toal, D., "In Search of Better E-Discovery Methods," *New York Law Journal*, April 23, 2008, available at <http://www.law.com/jsp/legaltechnology/pubArticleLT.jsp?id=900005509469>.

Buckley, C., Dimmick, D., Soboroff, I and Voorhees, E., "Bias and the Limits of Pooling," in *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 619-620, Seattle, 2006

Buckley, C. and Voorhees, E. "Retrieval System Evaluation," in *TREC: Experiment and Evaluation in Information Retrieval*, E. M. Voorhees and D. K. Harman, eds., MIT Press, pp. 53-75, 2002.

Cochran, W. *Sampling Techniques*, 3rd edition. John Wiley & Sons, New York, 1977.

Coleman v. Morgan Stanley, 2005 WL 679071 (Fla. Cir. Ct. Mar. 1, 2005)

Collaborative Expedition Workshop #45, Advancing Information Sharing, Access, Discovery and Assimilation of Diverse Digital Collections Governed by Heterogeneous Sensitivities, held Nov. 8, 2005, see http://colab.cim3.net/cgi-bin/wiki.pl?AdvancingInformationSharing_DiverseDigitalCollections_HeterogeneousSensitivities_11_08_05

Disability Rights Council of Greater Washington, et al. v. Washington Metropolitan Transit Authority, 242 F.R.D. 139 (D.D.C. 2007)

First International Workshop on Supporting Search and Sensemaking for Electronically Stored Information in Discovery Proceedings, available at <http://www.umiacs.umd.edu/~oard/desi-ws/>

ICAAIL 2007 (11th Annual Artificial Intelligence and the Law Conference) (selected readings from June 2007 Search and Retrieval Workshop known as "DESI I")

Jenkins, Tom, *Enterprise Content Management* (2005)

Mazza, M.; Quesada, E.; Stenberg, A., "In Pursuit of FRCP 1: Creative Approaches to Cutting and Shifting Costs of Discovery of Electronically Stored Information," 13 RICH. J.L. & TECH. 11 (2007), <http://law.richmond.edu/jolt/v13i3/article11.pdf>

Qualcomm Inc. v. Broadcom Corp., 2007 WL 2296441 (S.D. Cal.)

Dertouzos, J.; Pace, N.; Anderson, R., "The Legal and Economic Implications of E-Discovery" (2007 RAND Corp. Occasional Paper), available at http://www.rand.org/pubs/occasional_papers/OP183/

Schmidt, H.; Butter, K.; and Rider, C. "Building Digital Tobacco Document Libraries at the University of California, San Francisco Library/Center for Knowledge Management," *D-Lib Magazine*, 8(2), 2002.

Second International Workshop on Supporting Search and Sensemaking for Electronically Stored Information in Discovery Proceedings, available at <http://www.cs.ucl.ac.uk/staff/S.Attfield/desi/>

Sedona Best Practices Commentary on the Use of Search and Information Retrieval

Methods in E-Discovery (August 2007 public draft),
http://www.thesedonaconference.org/content/miscFiles/publications_html

Sedona Conference, *The Sedona Principles, Second Edition: Best Practices Recommendations & Principles for Addressing Electronic Document Production* (2007),
http://www.thesedonaconference.org/content/miscFiles/publications_html

Tomlinson, S., Oard, D.W., Baron, J.R., Thompson, P., “Overview of the TREC 2007 Legal Track,” *The Sixteenth Text Retrieval Conference (TREC 2007) Proceedings*, Gaithersburg, MD, November, 2007. <http://trec.nist.gov>.

Turtle, H., “Natural Language vs. Boolean Query Evaluation: A Comparison of Retrieval Performance,” *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 212-220, Dublin, 1994.

United States v. O’Keefe, 537 F. Supp. 2d. 14 (D.D.C. 2008)

U.S. Federal Rules of Civil Procedure, Rules 26 & 34, as amended (Dec. 1, 2006)

Victor Stanley, Inc. v. Creative Pipe, Inc., 2008 WL 2221841 (D. Md.)

Voorhees, E., “Variations in Relevance Judgments and the Measurement of Retrieval Effectiveness,” *Information Processing and Management*, 36(5)697-716, 2000.

Zubulake v. UBS Warburg, 217 F.R.D. 309 (S.D.N.Y. 2004)