AC 2011-624: TEACHING WEB INFORMATION RETRIEVAL AND NETWORK COMMUNICATION TECHNOLOGY TO NON-MAJOR UNDERGRADUATE STUDENTS

Xiannong Meng, Bucknell University

XIANNONG MENG is a Professor of Computer Science in the Department of Computer Science at Bucknell University in Lewisburg, Pennsylvania, U.S.A. His research interests include distributed computing, data mining, intelligent Web search, operating systems and computer networks. He received his Ph.D. in computer science from Worcester Polytechnic Institute in Worcester, Massachusetts, U.S.A.

Song Xing, California State University, Los Angeles

Song Xing received his B.S. and M.S. degrees in electrical engineering from Southeast University, China, in 1985 and 1990, respectively, and his Ph.D. degree in electrical & computer engineering from George Mason University, Virginia, in 2003.

From 1985 to 1995, he was a Lecturer in the Radio Engineering Department at Southeast University, China. He was also a Researcher at the National Mobile Communications Research Laboratory, China, from 1990 to 1995.

He was a Visiting Researcher in the Electrical & Computer Engineering departments at the University of Michigan-Dearborn from February 1995 to April 1995 and at Boston University from May 1995 to August 1996, respectively. From August 1997 to May 2003, he was an Instructor with the Electrical & Computer Engineering Department and Teaching/Research assistants in Electrical & Computer Engineering, Computer Science and Information Systems departments at George Mason University.

In 2003, Dr. Xing joined California State University, Los Angeles where he is currently an Associate Professor in the Information Systems Department.

His research interests include Internet statistical measurement, communication networks and digital systems, importance sampling simulations of stochastic systems, and speech and image processing.

Teaching Web Information Retrieval and Network Communication Technology to Non-Major Undergraduate Students

Abstract

World Wide Web and the Internet have touched our life in many aspects. From a technology point of view, both the Web and the Internet have been experiencing dramatic changes over the years. From a societal impact point of view, the Web and the Internet have become an essential tool in our live, similar to reading and writing, to a level where we have to study carefully the effect of this tool in our learning and in our live. This paper describes authors' experiences in teaching the fundamentals of World Wide Web search engines and the Internet technology to non-major undergraduate students. In one course, the first author taught a foundation seminar on the subject of web search engines to a group of first-year students from various disciplines from social sciences, humanity, natural sciences, and engineering. In the second course, the second author taught an upper level course on the subject of wireless and personal communications to non-majors from the business school. Both courses gave us rich experiences on how to reach non-major undergraduate students with topics of great importance and great interest.

World Wide Web and the Internet have touched our life in many aspects. From a technology point of view, both the Web and the Internet have been experiencing dramatic changes since the inception of the Internet in the late 60s and early 70s and the Web in the 90s. The technology of the Internet and the Web evolved from the ones used national laboratories for exclusive research projects with expensive equipment in the early days to what ordinary people today use for their daily life on devices such as cell phones and other consumer-oriented, small mobile devices. More and more people use the Web and the Internet everyday and more are curious about how the Web and Internet work. From a societal impact point of view, the Web and the Internet have become an essential tool in our live, similar to reading and writing, having reached a level where we have to study carefully the effect of this tool on our learning and on our live.

Traditionally, the subject of communications network and Web is taught in the context of science and technology, typically in a junior or senior course offered for majors in computer science, computer science and engineering, information technology, and other similar majors. These courses teach students the scientific principles and engineering technology involved in the design, implementation, testing, maintaining, and applications of the communications network and the Web. The students taking these courses typically have substantial computer science background such as programming, algorithms, data structures, operating systems, as well as proper mathematics background at or beyond calculus. In short, these courses were not accessible for students outside the aforementioned specialized majors. With the fast development and successful commercialization of technology such as mobile phone and e-commerce, the Internet and Web technology has reached ordinary people who do not have sophisticated technical background. The interest is growing fast of knowing how these basic technology work and how they may affect our live. How can we effectively teach the relevant technology and its societal impact to college students who have no technology background, or otherwise no interest

in the technology detail itself? It is in this context that the two authors would like to share with the community their experiences of teaching such courses.

This paper describes authors' experiences in teaching the fundamentals of World Wide Web search engines and the communications network technology to non-major undergraduate students. By non-majors we mean students are from outside the general disciplines of computer science or information technology. The two authors have taught two separate courses in two different universities in the United States. But the motivation, the course setting, and the experiences are very similar. And the challenges are very similar: how we can teach the Web search engines and communications network technology in general to undergraduate students who do not have previous computing background. The paper will concentrate on one course at Bucknell University in great detail first. (Course A.) The second course in California State University at Los Angels will be briefly described so the similar information will not be redundant. We will first describe the background, content, and structure for Course A. We then will discuss the pre and post course expectations. The assessment of the course will be discussed next, followed by a discussion of challenges and how we addressed them. We then will discuss the course at California State University – Los Angels (Course B), in which we will point out some similarities and differences between the two courses. Last we will conclude our discussion, drawing some lessons from our experiences and proposing future course of actions.

A Foundation Seminar at Bucknell University: Search Engines and Our Lives

Bucknell University requires its Arts and Science students to take a foundation seminar course during their first year (mostly in their first semester). Some Engineering students elect to take the seminar as well since it is very beneficial to the first-year engineering students. Here is a paraphrase quote from the university's course catalog about the foundation seminars[2].

> Each first-year student in the College of Arts and Sciences enrolls in a small seminar of about 15 students, usually in the fall semester. Foundation Seminars are offered by many different faculty and focus on a wide variety of subjects. Whatever the topics, they are designed to cultivate the attitudes, skills, and knowledge necessary for students to benefit maximally from a Bucknell University education and to negotiate the complexities of the modern world. The seminars stress the following: active, independent learning; collaborative learning; development of students' capacity for analysis, reflection, judgment, and creativity; multiple perspectives; and development of skills students need in order to engage in intellectual endeavors at Bucknell and beyond. These courses address foundation skills in reading, writing, listening, and speaking and also develop students' ability to use the library effectively and to use computers (e.g. word processing, simulations, use of a database, or analysis of data).

While the general outcomes of a foundation seminar at Bucknell University are cultivating the capability of life-long learning through reading, writing, listening, presentation, and becoming information literate, the specific goals of this foundation seminar *FOUN 090-25: Search Engines and Our Lives* are to make students aware of the general technologies used in a typical web search engine, understand the advantages and limitations of using search engines, appreciate the societal implications brought in by the search engines. Through the exercise of reading, writing,

presentation, and literature search in the seminar, students will become a better, more independent scholar after the seminar who is also knowledgeable about search engines.

The foundation seminars target first year students from any majors on campus who may have no background in computer science. The learning outcomes of the foundation seminar, according to the university's catalog, are as follows.

1. Students will develop writing, reading, speaking, listening, and information literacy skills necessary for collegiate-level academic work.
2. Students will develop capacities for independent academic work and become more accountable for their own learning.

We realize these outcomes by asking students to read papers, find extra references, synthesize what they read, write research papers, and present to the class of their findings. Because the seminar is intended for first-year non-major students in their first semester, the mathematics and computer science components from a typical web information retrieval course are removed from this seminar. Rather the seminar concentrates on general ideas of information retrieval and web search engines. For example, instead of studying detailed algorithms and data structures for inverted index systems, an essential component of any search engine, we just illustrate the ideas of inverted indexing using diagrams and explain how they work in an information retrieval system. We also spent a half of the semester on societal impacts of search engines investigating human side of the issues related to web search such as politics, health care, privacy, environment, e-commerce, academics, among others.

We use the book *Search Engine Society* by Alexander Halavais[4] as our main reference, accompanied by a number of other articles from research or popular publications. Alexander Halavais is a professor of communications at Quinnipiac University. His book contains a wealth collection of information about search engines and their social implications intended for general public. The book is suitable for first-year students of any major without background knowledge in computer science or information retrieval. The topics discussed in the seminar mainly follow the order of the book. We started with the basic building blocks of a search engine, followed by a brief history of search engines. We then presented a couple of examples of search engines, namely, AltaVista and Google, using research papers available to the public. The seminar then discussed various topics of search engine impacts on our lives listed at the end of previous paragraph. Instead of programming, students read, write, and present papers as they progress through the semester. Each week, students are asked to write a weekly reading journal based on the reading materials which can be a chapter from Halavairs' book, a conference paper, or a journal paper. Some chapters of the book edited by Amanda Spink and Michael Zimmer[9] are also used as reading materials. Students then are asked to write and present two papers of their own, one in the subject of search engine technologies, the other in the subject of societal impact of these technologies.

One of the goals of the seminar is to introduce the basic working principles of a typical search engine to non-majors without involving math and programming. A search engine is abstracted as a working system with four major components, indexing, ranking, crawling, and user interface. When discussing indexing, the basic structure of inverted indexing system was presented (see Figure 1) so students understand for a given query, how a list of relevant documents can be generated. We also briefly discussed parsing of a text file. Students were asked to complete an

assignment on paper to parse a given set of web pages into a collection of tokens and to build an inverted indexing system out of them. When discussing ranking, we talked about the elements that typically go into the ranking system such as page popularity using PageRank[1], importance of the query words (location, frequency, and fonts of the query words in the document), and general ideas of term frequency and inverse document frequency. When discussing crawling, the subjects of traversing web pages using breadth-first, depth-first, or priority were studied. Students were asked to hand traverse sample web pages using breadth-first or depth-first algorithms. Also mentioned was the robot exclusion protocol so students understand that web crawlers are supposed to follow the protocol and respect web site owner's right. We studied a few robot.txt files on some popular websites such as www.cnn.com, www.abcnews.com, and www.ebay.com. When discussing user interface, we made a special point of client-server computing model where the browser is the client and the search engine is the server. We also discussed the fact that the server technically is able to, and many search engines do in fact, record user actions on the search engine. This discussion becomes a natural transition to next segment of the seminar when we discuss user search behaviors and the issue of user privacy.

Throughout this segment of the seminar, we avoided programming and math details when search engine techniques were presented, concentrating only on the basic ideas. Students were able to grasp the ideas with written exercises such as traversing the web pages (graph), and computing PageRanks in its simple form (PageRank of a web page is the sum of PageRanks of other web pages pointing to it). Students were also able to construct correctly a basic inverted index system after applying parsing to a set of given web pages.
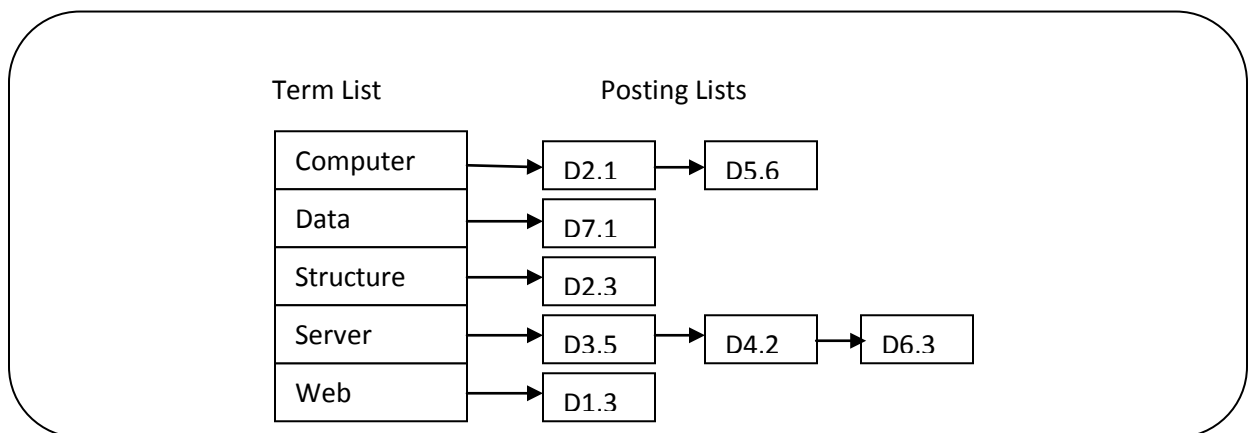

Figure 1: An Illustration of a Partial Inverted Indexing System

In the next segment of the seminar, we discussed search behaviors, that is, how users search the web and what kind of queries they have used in the past 10-15 years, using literature mostly from the research results of Spink and Jansen[8,5,9]. It was very natural for students to ask where the research data (user search queries) used in these research papers came from. The research data from Spink and Jansen's research came from search logs of a number of major search

engines over the years. Students were surprised that search engines were able (and allowed) to log so much detailed information about web search and its users. We also discussed the incident in which AOL released search logs of about 658,000 users in July 2006[6], which raises serious privacy concerns for search users.

After gaining a basic understanding of how search engines work, the seminar turned its attention to the social impact brought in by the search engine technology. We examined the effect of search engines in areas such as e-commerce, politics (presidential election and other political topics such as censorship in various countries) and cultural issues (different countries may have different views on the issue of freedom of speech on the internet, for example), environment issues (search engines use tremendous amount of energy and computer hardware in their data centers which result in environmental concerns), health care, and inequality in search among different segments of the society (see a list of reference papers used in the seminar from the course website[2]).

While a course project in a typical web search engine course for computer science majors would be a team programming project to build a simple search engine, the main student wok in this foundation seminar consists of various reading and writing assignments. The writing assignments helped students reflect on what they learn in the seminar. The assignments also are used to meet the writing goal for the foundation seminar. Two types of writing assignment are used in the seminar. One is the weekly reading journal. Students are asked to write a reading journal of 300 to 500 words every week on the subject of what was being discussed during that week. The other type of writing assignment is the research papers. Each student wrote two research papers during the semester, one written in team on the subject of search engine technology and its history which was due about half-way through the semester; the other written as an individual paper on the subject of social impact of search engines which was due at the end of the semester. The research papers are typically 1,200 to 2,500 words in length.

Students have some freedom in choosing the topics for their weekly reading journal. Most students follow the seminar schedule to write about the current chapter that is being discussed in the class, while others choose to write about what they read from the research papers suggested by the instructor outside the main reference text. The reading journals are required to have two segments, one summarizing the content of reading, the other reflecting on the reading. The instructor read all the journals each week and provided feedback to the students. Both the students and the instructor found this a great learning tool for a course of this kind. The reading journals serve the purpose of digesting the seminar content and reflecting on the discussions. We did not ask students to revise any of the journals. The two research papers by students, on the other hand, are required to have revisions. For the first paper, which is a team paper, students write about technology or history of search engines. Some wrote about crawlers, rankers, indexers, and user interfaces of search engines. Others wrote about history of search engines, or more general, history of information retrieval that goes all the way back to Bush's Memorex paper[3]. Student teams presented their paper to the whole class after finishing writing the paper. The second paper is by individual student. The students are asked to write about social implications of search engines. Again students have the freedom of choosing a specific topic. Students wrote about a wide range of issues such as social networking as advancement over search engines, academic plagiarism using search engines, medicines over the internet, privacy

in search engine era, among other very interesting topics. Each individual student was also required to make a public presentation about their research findings.

Although the seminar targets non-computer science first-year students, the subject of search engines inevitably touches many computer science topics. Whenever appropriate, we have a "computer science moment" while discussing search engine technology and its societal impact. The following are some examples of these "computer science moments".

- **Sorting**: We discussed that after computing the rankings of the relevant web pages, the search engine needs to sort the results before sending them to the user for displaying. The exact sorting algorithm used by the search engines can be complicated. We demonstrated the basic idea of selection sorting algorithm. Students actually were required to do a homework problem to sort term list on paper after parsing a few sample web pages using the selection sort algorithm.
- **Search**: Obviously search engines need to search term list when a query is presented. While we may not know the exact search algorithm that the search engines use, it is important to understand that searching is absolutely necessary in search engines. We also emphasized that the time needed to complete a search can be quite different using different search algorithms. Students were asked to exercise on paper to find search results using the binary search algorithm.
- **Graph traversal**: When studying the process of crawling, we talked about how a crawler can proceed once it visited the very first web page. The crawler can follow other web pages that are linked to this starting page by a breadth-first traversing, a depth-first traversing, or a priority-first traversing. Students are asked to do an exercise in breadth-first traversal on a set of inter-connected web pages.
- **IP address for networked computers**: When discussing how search engines may log user search data, one of the pieces of information of which search engines keep track is the IP address of the user computer. We discussed the need for a numerical address for computers to communicate over the network as well as the need for symbolic host names of networked computers so human beings can use the address easily.

While most of the 16 students in the seminar are students who did not have any computer science background, a few did show interest in computer science, or programming. They asked to learn more details about programming a search engine. The instructor demonstrated in the class a simple search engine that was developed by the instructor and his students in one of the computer science courses. We even used this simple search engine to search the content of the seminar so students can find anything they want from the lecture notes, reference materials, assignment and others available from the course website through this experimental search engine.

Pre and post expectations of the foundation seminar

Since the seminar is designed for first-semester college students of all majors in the College of Arts and Sciences at Bucknell University, no special technical background regarding computer science or mathematics are required. What is expected is the background knowledge a typical first-semester college student should have, certain level of reading and writing skills, certain level of mathematics maturity (not necessarily college calculus since not all arts and sciences

students are required to take calculus in their first semester), as well as experiences in using some types of web search engines.

As indicated in the introductory section of the paper, Bucknell University established the following two Learning Outcomes for Foundation Seminars.

1. Students will develop writing, reading, speaking, listening, and information literacy skills necessary for collegiate-level academic work.
2. Students will develop capacities for independent academic work and become more accountable for their own learning.

To realize the Learning Outcomes, the seminar expects, as the result of the course, for students to understand the technological and societal aspects of search engines and their impact on our daily life. These goals are achieved through various learning activities such as lecturing, individual reading, small group discussions, and class presentations. Students are expected to develop library and information technology skills to research and summarize information needed to support the views they formulate regarding particular issues. Students are also expected to develop writing, reading, speaking, listening, and information literacy skills necessary for collegiate-level academic work.

Assessment

The assessment comes from two aspects, one is the observation and reflection of the instructor after teaching the course, the other comes from the course evaluation required of all courses by the university. After reading all papers and journals written by the students and having participated in the presentations by the students, the author felt the seminar achieved its objectives. Students understand the basic working of search engines and their impact on society and on our lives. Students also improved their skills of writing, reading, and public presentation over the semester. From the instructor's point of view, the seminar was a success.

The course evaluation for foundation seminars are conducted very similar to any other courses at Bucknell University except that additional questions about writing are asked in a separate evaluation form for the writing component of the course. Bucknell University course evaluation asks a number of standard questions and then allows students to write free format comments. The one question that reveals how students think about the course in overall is "I would recommend this course to other students interested in this subject." This question received a score of an average of 3.57 out of 5, not a very high one. But one has to take into consideration the fact that only one student out of 16 in this seminar is a computer science major and the fact that this is the first semester at college for all students. Students are for the first time filling out these course evaluation forms. Still, the foundation seminar students have many positive things to say about the seminar when they wrote their free-text comments part of the course evaluation, "I liked learning about the future of the web.", "I liked learning about the current issues about the internet.", "I liked the weekly journal entry because it helped me understand the reading while working on my writing.", "Journals helped improve my writing.", "the weekly journal helped me understand the reading while working on my writing.", "the round-table discussions were a good way to stimulate classroom discussion and excellent discourse."

Challenges and responses

When designing the seminar, the first challenge was how to organized the seminar content at an appropriate level for the first-semester, non-major college students. We wanted to convey the technical information used by search engines, but we cannot rely on, nor should we utilize contents that are too technical such as programming and complicated mathematics. We ended up choosing an appropriate textbook[4] that provides sufficient detailed technical information yet was written for non-technical readers. We also filled in quite amount of detailed content that are suitable for students. The seminar asked students to do in-class work and homework that doesn't require sophisticated mathematics and programming (see the examples discussed in the content section of this paper). In author's opinion, this approach worked very well.

The second challenge was how to incorporate writing, reading, information literacy, and presentation into the seminar, as required by the rules of the university. What the author did was to use multiple means for students to practice various aspects of the requirement. Students wrote frequently. They are required to write weekly reading journals of half a page to a page long that are checked by the instructor (no detailed comments or grading). They also wrote two research papers that require literature search. One paper was a team project to investigate a technology aspect of search engines; the other paper was an individual research project to write about societal impact of search engines. Both papers required a public presentation. Reading what the students wrote and observing their presentations, it is clear that this part of the seminar goals were met through these exercises.

Special Topic Elective at California State University at Los Angels: Wireless Communications and Networks

In this section, we describe the course and experience of the second author who taught a "Wireless Communications and Networks" course to non-majors.

Wireless technology has quickly become the newest networking technology that has hit the mainstream of communications systems. The purpose of this course is to provide a straightforward and broad survey of wireless voice and data network standards and technologies available today for personal and business communications. It is designed for business and information management students taking an entry-level wireless technology course or seeking better knowledge of wireless communications and networks, assuming the students having little or even no technical background.

The course is a major elective in Management of Information System at the business school and its prerequisite is *CIS 100 Business Computer Systems*, which is offered to all majors at the College of Business and Economics at California State University at Los Angles. CIS 100 provides students with computer system fundamentals, computer hardware and software concepts, and introduction to microcomputer software. Unlike other courses on wireless telecommunications systems and networks which provide a deeper understanding of the *operations of wireless technologies* used by professionals and technicians involved in a technical support area of mobile computing and wireless networking, this elective course emphasizes on

the understanding of the *wireless networking systems concepts and principles*. The course is designed based on the assumption that the students taking this course had a basic knowledge of networking most likely from their non-technical experience using a wired or wireless computer network from Bluetooth and Wi-Fi to 3G/4G and satellite broadband at home or their business environment.

For the student learning objectives presented above, this course provides a fundamental introduction to all wireless communication systems including wireless personal area networks (infrared, Bluetooth, cordless phone), wireless local area networks (Wi-Fi), wireless metropolitan area networks (fixed and mobile WiMAX), and wireless wide area networks (cellular wireless, satellite communications). The topics cover the wireless networking systems' architecture, standards, technologies, QoS (quality of service), security, and multimedia applications in business.

The course starts at the data and computer network communications overview, followed by a brief introduction to the data transmission techniques like RF (radio frequency) communications, signaling, modulation, and multiplexing. Although the course avoids the requirement of students for math, programming, and deep analysis of technologies, we believe the brief coverage of data transmission fundamentals will help to illustrate how they fit together in a modern wireless network. The course then majorly focuses on each wireless communication system listed above and ends at the wireless and personal communications applications in business which includes advantages of wireless technology, challenges of using wireless technology, wireless Internet access using Wi-Fi, WiMax, and 3G/4G, building a wireless infrastructure, radio frequency identification (RFID), wireless applications in medical and healthcare, industrial and commercial wireless applications, multimedia in wireless: audiovisual telephony, videoconferencing, broadcasting audio & video, etc.

The hands-on exercises on wireless LAN (local area network) devices, configuration and installation, troubleshooting, and maintain small and medium-sized wireless networks are assigned to students regularly accompanying to their homework assignments to help students establish the direct experience using wireless technologies and help them to better understand and master the subject matter of the topics. Students can even use their home network facilities to fulfill the hands-on projects. In addition, a group project is required of students, which builds upon and complementing the material covered in class.

We have chosen the text books with the writing style for non-major undergraduate students and which were tried by non-technical writing language. Such references include *Fundamentals of Wireless Networking* by Ron Price, McGraw-Hill/Irwin, *Business Data Networks and Telecommunications* by Raymond Panko, Prentice Hall, and *Wireless# Guide to Wireless Communications* by Jorge Olenewa and Mark Ciampa, Cengage Learning.

The course topics we selected are also of assistance to students who are looking to obtain the Wireless# (or CWTS, Certified Wireless Technology Specialist) entry-level certification and the Certified Wireless Network Administrator (CWNA) foundation level certification from Planet3 Wireless, the organization that is the leader in vendor-neutral wireless certifications.

The College of Business and Economics at California State University at Los Angles has offered this course five times as a special topic since Spring 2007. The course has proved successful in student learning and attracted a large number of students each time from a variety of majors in the College of Business and Economics and other colleges in the university. The class was also very rewarding with high student evaluation scores each time.

Conclusions

The authors find the technical topics such as web search and communications network technology can be taught effectively to non-majors by removing some of the deep math and technical information from the courses. The authors in this paper described their experiences in teaching such courses in two different universities and for two different groups of students. *FOUN 090-25: Search Engines and Our Lives* taught in Bucknell University was a first-year foundation seminar where the author concentrated on the general topic of web search engines and their social impact. The second author taught the wireless communications and networks to business majors who do not have the same math and programming background as computer science majors. We removed the heavy-duty math and computer programming from the course. We added a large component of social impact of the web search and a component of writing and presentation. *CIS 454: Wireless and Personal Communications Systems* in California State University at Los Angles was about communications network technology. We taught the course in a way that was accessible to non-majors. Both authors and their students felt the courses were a great success, teaching students the basics of technology and having a deeper understanding of the social impact brought on by the latest technologies. Students also practiced their writing and presentation skills in these two courses.

References

[1] Brin, S. & Page, L. (1998). The Anatomy of a Large-Scale Hypertextual Web Search Engine. In *Proceedings of the 7th IWWW Conference*, Brisbane, Australia, 14-18 April 1998.

[2] Bucknell University. (2010). *Foundation Seminar*. Accessed December 20, 2010 from http://www.eg.bucknell.edu/~found-cs/2009-fall/index.html

[3] Bush, V. (July 1945). As We May Think. *The Atlantic Monthly*. July 1945.

[4] Halavais, A. (2009). *Search Engine Society* Malden, MA: Polity Press.

[5] Jansen, B.J. & Spink, A. (2006). How Are We Searching the World Wide Web? A Comparison of Nine Search Engine Transaction Logs. *Information Processing & Management*, 42(1), 248-263.

[6] Kawamoto, D. & Mills, E. (August 7, 2006). AOL apologizes for release of user search data. *CNET News*. Accessed December 29, 2009 from http://news.cnet.com/2100-1030_3-6102793.html

[7] Spink, A. & Jansen, B.J. (2008). Trends in Searching for Commerce Related Information on Web Search Engines. *Journal of electronic commerce research*. 9(2), 154-161.

[8] Spink, A., Jansen, B.J., Wolfram, D., & Saracevic, T. (2002). From e-sex to e-commerce: Web search changes. *Computer*, 35(3), 107 - 109.

[9] Spink, A. & Zimmer, M. (2008). *Web Search -- Multidisciplinary Perspectives*. Berlin Heidelberg: Springer-Verlag.