



INFORMATION SCIENCE RESEARCH Bringing Together the Disciplines to Focus on the Exploding Information Age

Many members of the Cornell community will remember 301 College Avenue as the Triangle Bookstore in Collegetown. Today, the same building bears a sign "Cornell Information Science." It is home to a new interdisciplinary program that studies information on computers and networks.

Research in Information Science concentrates on the areas where computer science and the social sciences overlap and reinforce each other. As digital technologies have become pervasive in our culture, researchers have found that technical, human, and social questions are interwoven and must be studied together. Information Science has a strong technical core, but it studies equally the human and social context in which information systems are employed. The faculty come from disciplines as varied as computer science, communication, science and technology studies, cognitive psychology, operations research, linguistics, economics, sociology, and law.

Several Cornell-based information systems are, themselves, subjects of research. They include the pioneering physics e-Print archive (arXiv.org), for which Paul H. Ginsparg, Computing and Information Science; Physics, received a MacArthur Fellowship in September 2002, and the Legal Information Institute created by Thomas R. Bruce and Peter W. Martin, Law, which is a focus for research in legal informatics and automated text processing. A fascinating system, which illustrates the diversity of digital information, is the Nuprl collection of mathematical theorems and proofs, which was built by a team under the leadership of Robert L. Constable, Computer Science.

The National Science Digital Library (NSDL) is a major National Science Foundation program to build a very large digital library for science education. William Y. Arms, Computer Science, and Carl J. Lagoze, digital library scientist, with colleagues at the University Corporation for Atmospheric Research and Columbia University, head a program of research to integrate heterogeneous digital libraries, such as those within the NSDL, into a coherent whole. Innovative applications such as these provide numerous opportunities for collaboration with the Cornell University Library.

One theme of Information Science is how people interact with information. An early impetus for forming the program was the close relationship between the Human Computer Interaction Group, based in the Communication Department, and digital libraries research in Computer Science. Members of the Human Computer Interaction Group include Geraldine K. Gay, Jeffrey T. Hancock, Helene A. Hembrooke, research associate, and Joseph B. Walther,

Communication; and Phoebe J. Sengers, Science and Technology Studies. The group investigates social, psychological, and design issues surrounding the use of computers at school, work, and home.

The Human Computer Interaction Group uses an iterative, user-centered design approach to build and evaluate networked, multimedia environments. The group has a human factors laboratory in the Information Science building. Their current projects focus on use of wireless technologies, information-seeking behavior, virtual groups, and social navigation. The group also provides a center of expertise in the methodology of social science experimentation, including educational evaluation, usability, and experimental design.

The theme of how humans interact with information through technology has created a fruitful partnership with the Cognitive Studies Program, directed by Shimon J. Edelman, Psychology. Many researchers are members of both programs. For example, Joseph Y. Halpern, Computer Science; Lawrence E. Blume and David A. Easley, Economics; and Michael W. Macy, Sociology, study the complex relationship between individual beliefs, decisions, and social reality. Edelman studies cognitive computational neuroscience of vision, and Michael J. Spivey, Psychology, uses psycholinguistic experimentation to study how humans extract semantics from information. Natural language processing and linguistics are further examples of Information Science research on the boundary between computer science and cognition.

A second theme that pervades much of Information Science research is the use of computing methods to automate tasks that require intellectual effort by skilled staff. The goal is to develop highly effective methods that operate on huge bodies of information, such as the Web, at costs that are far lower than conventional, labor-intensive methods. An example is the use of metadata—data about data, such as catalogs, indexes, abstracts, and classification schemes.

Traditionally, metadata has been generated by professional staff using procedures that are expensive and time-consuming, yet surprisingly, little is known of its usefulness. Jayavel Shanmugasundaram, Computer Science, and Lagoze explore metadata models that apply to the semi-structured data found on the Web, while Gay and colleagues from Syracuse University study how people actually use metadata. Donna Bergmark, program analyst specialist, combines methods from information retrieval with Web crawling to automate the selection of material without requiring metadata, for reading lists, as example.

The technical foundation of automated information retrieval is the vector space model for representing text, developed by Gerard Salton at Cornell in a classic series of experiments begun in the 1960s. This approach is entirely statistical. Computer scientists are now developing automated methods that go far beyond treating text as a sequence of separate words. Claire T. Cardie and Lillian J. Lee, Computer Science, use natural language processing to extract information from documents, including the inter-relations between documents, while Mats Rooth, Linguistics, applies computational linguistics to problems of semantic information in text. Richard A. Caruana and Thorsten Joachims, Computer Science, use machine learning to probe Web information. Joachims works with Ginsparg to analyze and evolve the physics e-Print arXiv automatically and to assist users in interacting with it.

Going beyond text, Johannes E. Gehrke, Computer Science, has developed some of the fastest methods available for automated data mining and privacy-preserving techniques that do not reveal the data.

The Human Computer Interaction group uses eye-tracking equipment to observe in detail how people explore information online. The Intelligent Information Systems Institute, led by Carla P. Gomes, research associate in Computer Science, carries out research in computationally and data-intensive methods for intelligent decision making.

These automated methods need a theoretical and scientific basis. Computer scientists, such as John E. Hopcroft, Computer Science; Daniel P. Huttenlocher, Johnson Graduate School of Management; Computer Science; and Jon M. Kleinberg and Bart Selman, Computer Science, are adding a qualitatively new level of rigor, experimental capability, and scope to Web research. Kleinberg was one of the first people to study the structure of the Web from a theoretical as well as from a pragmatic viewpoint, by analyzing the Web as a huge graph of documents connected by hyperlinks. This view of the Web as a static graph needs elaboration to account for many characteristics of the Web as it is today.

One area for research is the role of temporal phenomena in the Web—long-range trends, the daily evolution of Web content, the minute-by-minute dynamics of news coverage, and the second-by-second dynamics of usage data. Current research at Cornell looks at changes in content as measured by bursts of activity and the stability of information structures, using techniques drawn from cluster analysis. Kleinberg and Ginsparg are looking for bursts of activity in the Physics arXiv; a potential application is to identify research papers on fast-moving topics for accelerated peer review or to bypass peer review entirely.

Carla Terzizi



Information Science at 301 College Avenue

Robert Barker/CU



Johannes Gehrke (l.), Computer Science, and William Arms (r.), Director of Information Science.

Information Science research has a final theme; information exists in a social and cultural context. In addition to collaborative activities with economics, sociology, and law, the program has two joint appointments with the Department of Science and Technology Studies. Tarleton L. Gillespie studies the relationship between society, technology, and policy in areas such as digital copyright, while Phoebe J. Sengers analyzes assumptions about work life and consumer culture that are embedded in new technologies, using the results to build new multimedia devices derived from alternative assumptions.

Cornell's Information Science program is exceptionally strong in technology, has great potential in Human Computer Interaction and Cognition Studies, and is growing rapidly in the social aspects of the field. One final step remains: to establish a first-rate doctoral program. This spring, the university approved a new graduate field and Ph.D. program, subject to approval by New York State. With this in place, Cornell can truly aim to become the world's premier research program in Information Science.

William Y. Arms

Director of Information Science
Professor of Computer Science

For more information:

<http://www.cis.cornell.edu/infoscience/>

Cornell's Educational Programs in Information Science

Educating a New Generation of Students Who Thrive on Different Approaches

Digital technologies have become pervasive in culture, economy, law, government, and research, dramatically changing the way people work and live. To comprehend these developments, Information Science is developing programs that cut across the traditional divisions of the university, combining computer science with the social sciences of how people and society interact with information. The goal is to educate a new generation of undergraduate and Ph.D. students who are both computer scientists and social scientists, and who thrive on the different approaches to theory, investigation, design, and practice of the established disciplines.

With this goal in mind, a group of Cornell faculty spanning more than a dozen departments and three Cornell undergraduate colleges have spent the past two years answering the question: What kind of education should a great university provide to its students in this Information Age? In response, they created a collection of interdisciplinary educational programs, at both the undergraduate and graduate levels, that complement the ongoing research efforts in Information Science.

- An undergraduate concentration, or minor, in Information Science was approved in six of Cornell's seven undergraduate colleges during the spring and summer of 2002. Close to 40 students adopted the concentration in its first year of existence. As a confirmation of the broad appeal of the field, students pursuing an Information Science concentration emanate from four colleges—Arts and Sciences, Agriculture and Life Sciences, Engineering, and Industrial and Labor Relations—and from among them, are pursuing fourteen different undergraduate majors.
- A new undergraduate major in Information Science was approved by the faculties of Arts and Sciences and Agriculture and Life Sciences during the spring of 2003; the College of Engineering is expected to follow suit next fall. Pending approval by the New York State Education Department, Information Science plans to launch the majors for the 2004–05 academic year. Because the new major will exist in the three largest undergraduate colleges, it has the potential for dramatic impact on undergraduate education at Cornell.
- A graduate field in Information Science was approved by the Graduate School and the Cornell Faculty Senate in spring 2003. Subject to approval by New York State, plans for the graduate field include admission of its first class of Ph.D. students in fall 2004.

Cornell's educational programs in Information Science are united in the study of the design and use of information systems in a social context: courses in each of the programs study the creation, representation, organization, application, and analysis of information in digital form. The focus of Information Science at Cornell is on systems and their use, rather than on the computing and communication technologies that underlie and sustain them.

Cornell's Information Science programs examine the social, cultural, economic, historical, legal, and political contexts in which information systems are employed, both to inform the design of such systems and to understand their impact on individuals, social groups, and institutions. The field's interdisciplinary research combines multiple methodologies, including mathematical analysis, computer modeling, software system design, experimental studies, and critical social evaluations, from such traditional disciplines as computer science, cognitive psychology, social science, cultural studies, and history.

With the creation of these exciting and progressive degree programs, the faculty in Information Science hope to draw strong, technically-inclined students from across the campus to the courses. The Information Science faculty strives to establish Cornell as a world leader in Information Science.

Claire T. Cardie

Associate Professor of Computer Science