INDIANA UNIVERSITY

University Graduate School
2009-2010
Academic Bulletin

Informatics

School of Informatics and Computing Bloomington

Dean
Robert Schnabel*

Associate Dean for Graduate Studies and Research
Geoffrey C. Fox*

Associate Dean for Faculty Affairs
David Leake*

Director of Graduate Studies
Martin Siegel*

Director of Graduate Admissions
David Hakken*

Departmental E-mail
graduate@informatics.indiana.edu

Departmental URL
informatics.indiana.edu

Faculty
(An asterisk [*] denotes membership in the University Graduate School faculty with the endorsement to direct doctoral dissertations.)

Professors
J. Michael Dunn* (Emeritus), Geoffrey Fox*, David James Hakken*, Michael McRobbie*, Christine Ogan* (Emerita), Edward Robertson* (Emeritus), Martin Siegel*, Erik Stolterman*, Peter Todd*, Alessandro Vespignani*, Larry Steven Yaeger*

Associate Professors
Eli B. Blevis*, L. Jean Camp*, Mehmet M. Dalkilic*, Dennis P. Groth*, Esfan Haghverdi, Sun Kim*, Filippo Menczer*, John Paolillo*, Christopher S. Raphael*, Luis M. Rocha*

Assistant Professors
Jeffrey Bardzell*, Shaowen Bardzell, Alessandro Flammini*, Matthew Hahn*, Eden Medina*, Steve Myers*, Predrag Radivojac*, Kalpana Shankar*, Haixu Tang*, XiaoFeng Wang*, David Wild*, Yuqing (Melanie) Wu*, Yuzhen Ye

University Graduate School
Kirkwood Hall 111
Indiana University
Bloomington, IN 47405
(812) 855-8853
Contact: grdschl@indiana.edu

Degrees Offered

The Doctor of Philosophy (Ph.D.) degrees in Informatics, and Computer Science and, the Ph.D. Minor in Informatics, and the Ph.D. Minor in Bioinformatics are offered through the University Graduate School. In addition, the School of Informatics and Computing offers the Master of Science in Bioinformatics, the Master of Science in Chemical Informatics, the Master of Science in Computer Science, the Master of Science in Human-Computer Interaction Design, and the Master of Science in Security Informatics (see the School of Informatics and Computing graduate bulletin).

Ph.D. in Informatics

The Ph.D. in informatics provides a balance between technological, scientific, and social dimensions involved in the development and application of information technology.

Admission Requirements
Admission requirements in the areas of undergraduate grade point average and GRE score levels are those of the University Graduate School. The applicant must have some direct familiarity with computation.

For students planning to focus on bioinformatics or chemical informatics, a high level of computer programming competence is required. Students focusing in health informatics are expected to have a background in one of the health care professions. Students planning to specialize in social informatics or human-computer interaction should have familiarity with design principles and have some grounding in the social sciences.

For those who enter the Ph.D. program directly from their bachelor's program, there will be a formal assessment after two years of coursework, an "up or out" evaluation. Assessment will look at successful progression in the Ph.D. program with regards to progress toward completion of course requirements, maintenance of course grades and overall GPA according to Graduate School guidelines, and research, as measured by presentations at disciplinary meetings and publications.
For those who wish to enter the Ph.D. program from their master's program, there will be an application process. In this case, there is a natural evaluation of the student's record. Upon matriculation, an advisor, which may be temporary, will be assigned to the applicant. This advisor will help guide the student to his or her intended focus until a full-time advisor is found.

Annual Review
Each year, students will be required to file an annual review with their advisor and program or dissertation committee. The review covers the period of the previous academic year and is

due June 1. Four areas will be covered: coursework, research, teaching, and service. Written feedback will be provided by the student's advisor.

Course Requirements
A total of 90 credit hours are required. There are 27 required credits, which include I501, 6 credits of seminar work, 9 credits in 3 different subdisciplinary or 'breadth' areas, also called 'core' courses, 3 credits in professionalism and pedagogy (most likely taken as I600), and 6 credits of research rotation (I790). Students must take an additional 12 credits of theory and methodology courses applicable to the student's specialty. These courses can be taken inside or outside the school. Students must also take an additional 21 to 30 credits in elective coursework. The required Ph.D. minor is included in this category. The remaining 21 to 30 credits will be taken in dissertation credits.

No more than 30 hours may be counted from a master's degree taken at Indiana University or a graduate program at another university. (An additional 6 hours of master's thesis or capstone project may be counted toward the Ph.D. at the discretion of the student's program committee, assuming the thesis or capstone project is of sufficient research quality.)

Tracks of Study
Choices of fields offered for qualifying examinations must be approved by the Graduate Studies Committee. Tracks of study currently proposed within the department are bioinformatics, chemical informatics, human-computer interaction, social informatics, complex systems, music informatics, security informatics, logic and mathematical foundations, and data discovery.

Minor
All students are required to have an appropriate minor approved by the University Graduate School. Minors will be selected with the advisor's recommendation. Some of the courses included in the minor may also count towards the student's methodology or other requirements.

Grades
An overall B (3.0) average for all Ph.D. courses in Informatics is required. A student whose cumulative grade point average falls below 3.0 for two consecutive semesters is subject to dismissal from the program.

Written Qualifying Examinations
All students will take a written qualifying examination that consists of a depth exam and a breadth exam. The qualifying examinations are described in a separate document. Examinations will be offered at the end of August and at the beginning of the second semester in January. Examinations must be completed by the beginning of the student's fourth year in the program but can be completed before that time when the core courses are completed. Students who do not successfully complete the examination can retake the exam a second time.

Oral Qualifying Examination
The oral qualifying examination covers in-depth knowledge of the student's primary research area. This examination is administered by the student's program committee. The qualifying examinations will normally be completed at the end of course work, before the student embarks on the dissertation. The student must pass this examination before passing on to candidacy.

Dissertation Proposal
The proposed research for the dissertation must be approved by the research committee and presented at a public colloquium in the school.

Final Examination
Oral defense of the dissertation.

Ph.D. Minor in Informatics

A minor in informatics requires 9 credit hours. The required 9 credit hours refer to any 3 graduate courses suitable for the student's research, to be decided by the student's advisor (in his or her department) and the Informatics graduate program director. Typically, these 3 graduate courses are chosen from the set of core courses available in the Informatics Ph.D. program.

Ph.D. Minor in Bioinformatics

Bioinformatics draws on knowledge and information from various fields such as biology, computer science, medicine, chemistry and physics. Students in relevant Ph.D. programs such as biochemistry and molecular biology, medical and molecular genetics, medicine, chemistry, or biology are the target audience for the Ph.D. minor in bioinformatics.

Requirements
A minor in bioinformatics requires 12 credit hours. The core curriculum consists of graduate level courses in informatics. Electives may be chosen based on personal interests from a broad list of courses in biology, chemistry, computer science, information science, and medical and molecular genetics. The graduate bioinformatics courses in the School of Informatics and Computing assume a minimal knowledge of cell and molecular biology. That level of understanding could be gained with at least 6 undergraduate credit hours in molecular biology, genetics, or evolution.

Courses

Core Courses
I601 Introduction to Complex System (3 cr.) P: MATH M118, INFO I201, or equivalent course. The course will cover fractals, emergent behavior, chaos theory, cooperative phenomena, and complex networks. Students will learn how to think differently about complex realities, finding ways to understand their complexity and addressing the problems they pose.

I604 Human Computer Interaction Design Theory (3 cr.) The course will explore, analyze and criticize underlying assump-

tions and the rationale behind some of the most influential theoretical attempts in HCI and related fields. The purpose of the course is to make students aware of how theories can influence practice and to develop critical thinking around the role, purpose, and function for theories.

I605 Social Foundations of Informatics (3 cr.) Topics include the economics of information businesses and information societies, legal and regulatory factors that shape information and information technology use, the relationship between organization cultures and their use of information and information technology, and ownership of intellectual property.

I611 Mathematical and Logical Foundations of Informatics (3 cr.) P: Basic Discrete Mathematics equivalent to MATH M 118, or consent of instructor. An introduction to mathematical methods for information modeling, analysis and manipulation. Topics include proof methods in mathematics, models of computation, counting techniques and discrete probability, optimization, statistical inference and more advanced topics that include but are not limited to Markov chains and random walks, random graphs, and Fourier analysis.

I617 Informatics In Life Sciences and Chemistry (3 cr.) Introduces the fundamental notions in genome and proteome informatics and chemical informatics, focusing on the design and organizing issues in information systems used in those areas. The course is designed for students with no biology or chemistry background, but some knowledge in informatics, who want to learn basic topics in bioinformatics and chemical informatics.

I651 The Ethnography of Information (3 cr.) *This course is currently not being offered.

Other Courses
G901 Advanced Research (6 cr.)

I500 Fundamental Computer Concepts for Informatics (3 cr.) An Introduction to fundamental principles of computer concepts for Informatics study, including an overview of computer architecture, computer algorithms, fundamentals of operating systems, data structure, file organization and database concepts. INFO I500 is expected to impart the required level of competency in computer science. This course may be waived in lieu of six undergraduate credit hours of computer science or informatics coursework, covering areas of programming, discrete structures, and data structures

I501 Introduction to Informatics (3 cr.) P: Graduate standing. Basic information representation and processing; searching and organization, evaluation and analysis of information. Internet-based information access tools; ethics and economics of information sharing.

I502 Informatics Management (3 cr.) P: Computer science course 300 level or higher. Survey of data management issues in medical, health, chemical and biology related areas; basic techniques of physical database structures and models, data access strategies, management and indexing of massively large files.

I504 Social Dimensions of Science Informatics (3 cr.) Examines ethical, legal, and social issues surrounding contemporary research and practice in science informatics. Topics include the nature of science and technology, the ramifications of recent advances in science informatics, and relevant science policy and research ethics. General knowledge of science informatics is assumed.

I506 Globalization and Information (3 cr.) Explores the processes that promote and impede movement of human action and informational activities to the most general levels, e.g., the level of the world as a whole. Surveys diverse theories of globalization to identify the best approaches for professional informatics career planning and making information globally accessible.

I519 Introduction to Bioinformatics (3 cr.) P: One semester programming course or equivalent. Sequence alignment and assembly; RNA structure, protein and molecular modeling; genomics and protenomics; gene prediction; phylogenic analysis; information and machine learning; visual and graphical analysis bioinformatics; worldwide biologic databases; experimental design and data collection techniques; scientific and statistical data analysis; database and data mining methods; and network and Internet methods.

I520 Security for Networked Systems (3 cr.) This course is an extensive survey of system and network security. Course materials cover the threats to information confidentiality, integrity and availability and the defense mechanisms that control such threats. The course provides the foundation for more advanced security courses and hands-on experiences through course projects.

I521 Malware Epidemic: Threat and Defense (3 cr.) This course is designed to be research and hands-on oriented. Students are required to read and present research papers that reflect the state of the art in malware-related research and participate in course projects that expose them to the cutting-edge technologies on malware defense.

I525 Organizational Informatics and Economic Security (3 cr.) Security technologies make explicit organizational choices that allocate power. Security implementations allocate risk, determine authority, reify or alter relationships, and determine trust extended to organizational participants. The course begins with an introduction to relevant definitions (security, privacy, trust) and then moves to a series of timely case studies of security technologies.

I529 Machine Learning in Bioinformatics (3 cr.) P: INFO I519, or equivalent knowledge. The course covers advanced topics in Bioinformatics with a focus on machine learning. The course will review existing techniques such as hidden Markov models, artificial neural networks, decision trees, stochastic grammars, and kernel methods. Examine application of these techniques to current bioinformatics problems including: genome annotation and comparison, gene finding, RNA secondary structure prediction, protein structure prediction, gene expression analysis, proteonmics, and integrative functional genomics.

I530 Seminar in Health Informatics Applications (3 cr.) P: Graduate standing. This course examines the basic concepts of the design, evaluation, and use of interactive applications in health informatics.

I531 Seminar in Health Informatics (1-3 cr.) Variable topic. Emphasis is on advanced topics and research in health informatics. Can be repeated once with a different topic, subject to approval of the program director.

I532 Seminar in Bioinformatics (1-3 cr.) Variable topic. Emphasis is on advanced topics and research in bioinformatics. Can be repeated with different topics, subject to approval of the Dean.

I533 Systems & Protocol Security & Information Assurance (3 cr.) This course looks at systems and protocols, how to design threat models for them and how to use a large number of current security technologies and concepts to block specific vulnerabilities. Students will use a large number of systems and programming security tools in the laboratories.

I534 Seminar in Human-Computer Interaction (1-3 cr.) P: Graduate standing. Variable topic. Emphasis is on advanced topics and research in human-computer interaction. Can be repeated once with a different topic, subject to approval of the program director.

I536 Foundational Mathematics of Cybersecurity (3 cr.) P: Knowledge of undergraduate level probability, lined algebra or calculus. Students will learn mathematical tools necessary to understand modern cyber security. The course will cover introductory mathematical material from a number of disparate fields, including probability theory, computational theory, complexity theory, group theory, and information theory.

I537 Legal and Social Informatics of Security (3 cr.) This is a case-based course on privacy and security in social contexts. Cases will particularly address the specific designs of technologies (e.g., P3P, PICS) and discuss how different technically feasible design choices would result in distinct regulatory regimes, business strategies, or support different forms of social interaction. This course will focus on specific security and privacy technologies as socio-technical systems.

I537 Legal and Socials Informatics of Security (3 cr.) This is a case-based course in privacy and security in social contexts. Cases will particularly address the specific designs of technologies (e.g., P3P, PICS) and discuss how different technically feasible design choices would result in distinct regulatory regimes, business strategies, or support different forms of social interaction. This course will focus on specific security and privacy technologies as socio-technical systems.

I538 Introduction to Cryptography (3 cr.) Introduction to the foundational primitives of cryptography and implementations. A primary goal of this course will be to understand the security definitions for each primitive, and how they are used in cryptographic protocols. The ethics of insecure or on-the-fly protocol design will be discussed.

I539 Cryptographic Protocols (3 cr.) Provides a basic understanding of computer security by looking at how things go wrong and how people abuse the system. Once it is understood how computer systems are attacked, it is possible to propose ways to make the system secure.

I541 Interaction Design Practice (3 cr.) Human-Computer Interaction Design (HCID) describes the way a person or group accomplishes tasks with a computer - what the individual or group does and how the computer responds; what the computer does and how the individual or group responds. Sometimes known as “interface design,” HCID becomes increasingly important as computing intelligence and connectivity spread ubiquitously to home, work, and play environments. This course will be organized around a collection of readings and three design projects concerned with applying human-computer interaction principles to the design, selection, and evaluation of interactive systems.

I543 Interaction Design Methods (3 cr.) Students will learn basic concepts and methods for usability studies and evaluation of interactive systems as well as apply those methods to actual system design evaluations. This course is not only for understanding the basics and traditional approaches in this area, but also for exploring new ways of evaluating the usability of state-of-the-art technology-based systems such as systems in ubiquitous computing, CSCW, tangible and social computing areas.

I545 Music Information Representation, Search, and Retrieval (3 cr.) P: Major, minor, or outside area standing in music informatics or music information technology. A comprehensive, comparative study of computer-based representation schemes for music, including those oriented toward music notation, music performance, and music analysis. Overview of musical metadata. Techniques and tools for search and retrieval of music information. Credit not given for both INFO I545 and MUS N564.

I546 Music Information Processing: Symbolic (3 cr.) This course deals with both methodology and specific applications that attempt to algorithmically annotate, understand, recognize, and categorize music in symbolic (score like) form. Particular applications will include key finding, harmonic analysis, note spelling, rhythm recognition, meter induction, piano fingering, and various classification problems such as genre or composer identification. The methodology we will employ will be probabilistic and will include ideas from Machine Learning such as optimal classifiers, hidden Markov models, and Bayesian Networks. Students will have computing assignments, present papers, and be expected to implement solutions to problems using a high-level language such as R or Matlab.

I547 Music Information Processing: Audio (3 cr.) This course deals with various music analysis and processing problems that use sampled audio as the primary data representation. Discusses digital signal processing, including filtering and its relationship to Fourier techniques. Topics include synthesis, effects processing, score following, blind music recognition, and

accompaniment systems.

I548 Introduction to Music Informatics (3 cr.) P: Solid understanding of music fundamentals; music theory background recommended. History, issues, and applications in music information technology. Survey of various types of musical information. Introduction to digital musical media, including data standards and processing; database structure and organization standards and processing; database structure and organization of audio-, score-, and text file objects; and discussion of copyright issues.

I571 Chemical Information Technology (3 cr.) P: Consent of instructor**.** Overview of chemical informatics techniques, including chemical structure coding, chemical data representation, chemical database and search systems, molecular visualization and modeling techniques, and the development of chemical informatics software.

I572 Computational Chemistry and Molecular Modeling (3 cr.) P: INFO I571. Computer models of molecules and their behavior in gas and condensed phases; implicit and explicit solvation models; quantum and molecular mechanics; search strategies for conformational analysis, geometry optimization methods; information content from Monte Carlo and molecular dynamics simulations; QSAR; CoMFA; docking.

I573 Programming for Science Informatics (3 cr.) Students will receive a thorough understanding of software development for chem- and bioinformatics, and broaden experience of working in a scientific computing group. Topics include programming for the web, depiction of chemical and biological structures in 2D and 3D, science informatics tool kits, software APIS, AI and machine-learning algorithm development, high performance computing, database management, managing a small software development group, and design and usability of science informatics software.

I585 Bioinspired Computing (3 cr.) Biologically-inspired computing is an interdisciplinary field devoted to computational methods modeled after natural design principles. The goal is to produce informatics tools with enhanced robustness, scalability, flexibility and natural human-machine interaction. Topics include: Self-organization, Evolutionary Systems, Cellular Automata, Boolean Networks, L-Systems, Collective and Swarm Behavior, Artificial Immune Systems, Complex Networks.

I586 Artificial Life (3 cr.) Artificial life is a broad discipline encompassing the origins, modeling, and synthesis of natural and artificial living entities and systems. Artificial intelligence, as a discipline, tries to model and understand intelligent systems and behavior, typically at the human level.

I590 Topics in Informatics (1-3 cr.) P: Graduate standing. Variable topic. Emphasis is on new developments and research in informatics. Can be repeated with different topics, subject to approval of the Associate Dean for Graduate Studies.

I591 Graduate Internship (0-6 cr.) P: Approval required. Students gain professional work experience in an industry or research organization setting, using skills and knowledge acquired in Informatics course work. May be repeated for a maximum of 6 credit hours.

I602 Music Information Processing: Audio (3 cr.) This course deals with various music analysis and processing problems that use sampled audio as the primary data representation. Digital signal processing including filtering and its relationship to Fourier techniques. Focus on applications including score following, automatic music transcription and annotation from audio,

musical accompaniment systems, as well as some useful audio effects.

I609 Advanced Seminar I in Informatics (3 cr.) P: Advanced graduate standing or consent of instructor. Ph.D. student introduction to major historical and emerging theories, methods, technologies, and applications in Informatics. Provides students with opportunities to explore relevant research literature, results, and applications. Students will develop a profound understanding of leading research approaches and paradigms in their research area.

I619 Structural Bioinformatics (3 cr.) Informatics approaches addressing the sequence and 3D structure of biological macromolecules (DNA, RNA, Protein), with the objective of improving understanding of the function of these molecules. Topics will include molecular visualization; structure determination, alignment, and databases; and prediction of protein structure, interactions, and function.

I621 Computational Techniques in Comparative Genomics (3 cr.) Summarizes computational techniques for comparing genomes on the DNA and protein sequence levels. Topics include state of the art computational techniques and their applications: understanding of hereditary diseases and cancer, genetic mobile elements, genome rearrangements, genome evolution, and the identification of potential drug targets in microbial genomes.

I624 Advanced Seminar I in Human-Computer Interaction (3 cr.)
I625 Advanced Seminar I in Social Informatics (3 cr.)
I627 Advanced Seminar I in Bioinformatics (3 cr.)
I628 Advanced Seminar I in Complex Systems (3 cr.)
I634 Advanced Seminar II in Human-Computer Interaction (3 cr.)
I635 Advanced Seminar II in Social Informatics (3 cr.)
I637 Advanced Seminar II in Bioinformatics (3 cr.)
I638 Advanced Seminar II in Complex Systems (3 cr.)
I647 Advanced Seminar I in Chemical Informatics (3 cr.)
I657 Advanced Seminar II in Chemical Informatics (3 cr.)

I667 Seminar in Health Informatics II (3 cr.) P: INFO I530. Advanced graduate seminar in health informatics, designed to complement INFO I530. This seminar is intended for graduate students enrolled in the Health Informatics track in the Informatics Doctoral Program.

I690 Topics in Informatics (1-3 cr.) P: Graduate standing. Variable topic. Emphasis is on new developments and research in informatics. Can be repeated with different topics, subject to approval of the Dean. Course is intended for Ph.D. students in the School of Informatics.

I698 Research in Informatics (1-12 cr.) Research not dissertation related under the direction of a member of the graduate faculty. May be repeated for credit for a total of 30 credit hours.

I699 Independent Study in Informatics (1-3 cr.) P: Consent of instructor. Independent readings and research for Ph.D. students under the direction of a faculty member, culminating in a written report. May be repeated for a maximum of 12 credit hours.

I709 Advanced Seminar II in Informatics (3 cr.) P: Advanced graduate standing or consent of instructor. Ph.D. student introduction to major historical and emerging theories, methods, technologies, and applications in Informatics and its sub-areas. Provides students with opportunities to explore relevant research literature, results, and applications. Seminar II, unlike Seminar I, focuses on recent advances in sub-areas of Informatics.

I790 Informatics Research Rotation (3 cr.) Working with faculty to investigate research opportunities. May be repeated for a total of 6 credit hours.

I798 Professional Practicum/Internship (non-credit) P: Current enrollment in graduate degree program in Informatics. Participation in graduate level professional training and internship experience.

I890 Thesis Readings and Research (1-12 cr.) Research under the direction of a member of the graduate faculty leading to a Ph.D. dissertation. May be repeated for credit for a total of 30 credit hours.