Expert Speakers

Friday, May 28, 10am, 2405 Siebel Center

“BeeSpace: An Interactive System for Functional Analysis Using Semantic Indexing of Biological Literature”

Dr. Bruce Schatz, Department of Medical Information Science, University of Illinois at Urbana-Champaign (http://www.beespace.uiuc.edu)

BeeSpace is the flagship bioinformatics project in the NSF Frontiers of Integrative Biological Research program. Supporting new computer science research, we have developed an interactive system for functional analysis.

This uses language processing for information extraction on biological literature, particularly for the new paradigm of automatic extraction on special collections. Semantic indexes are built to support conceptual navigation of gene function. Generating these indexes requires extensive computation on parallel computers and cloud clusters, including being the largest internal user of the experimental cloud cluster. The special collections can be selected via realtime clustering of document collections or via steerable clustering with specified topics. The navigation can be supported via document summarization of gene functions, using an entity recognition variant with biology instead of newspaper entities. Direct functional analysis is supported by computing the concept lists of the most discriminating terms within the collection corresponding to a list of genes. All the services can be interactively combined during a session to locate a region of the information space containing the most useful functional information. This is a practical bioinformatics application of computer science research into natural language processing.

(Slides from this talk are available here )
Friday, June 4, 1pm, 2405 Siebel Center

"From Information to Knowledge: Harvesting Entities and Relationships from Web Sources"

Dr. Gerhard Weikum, Max-Planck-Institut für Informatik (http://www.mpi-inf.mpg.de/~weikum/)

There are major trends to advance the functionality of search engines to a more expressive semantic level. This is enabled by the advent of knowledge-sharing communities such as Wikipedia and the progress in automatically extracting entities and relationships from semistructured as well as natural-language Web sources. Recent endeavors of this kind include DBpedia, EntityCube, KnowItAll, ReadTheWeb, and our own YAGO-NAGA project (and others). The goal is to automatically construct and maintain a comprehensive knowledge base of facts about named entities, their semantic classes, and their mutual relations as well as temporal contexts, with high precision and high recall. This talk discusses state-of-the-art methods, research opportunities, and open challenges along this avenue of knowledge harvesting.

Wednesday, June 9, 1pm, 2405 Siebel Center

"Finding and understanding multiple objects in images and videos"

Dr. Deva Ramanan, University of California-Irvine

Object recognition is one of the fundamental problems in computer vision. Classic approaches tend to focus on simplified scenarios of images containing a single object. Real-world images contain many objects defined by particular spatial interactions (people tend to ride atop bikes rather than under them). In this talk, I will focus on algorithms for finding and understanding multiple objects in images and videos. I will survey a variety of work done in our lab, from low-level tasks such as multi-object tracking, extracting the relative depth ordering of objects in the scene, to high-level tasks such as understanding actions defined by multi-object interactions. From a learning and inference point of view, such problems involve the novel application of tools such as structured prediction from machine learning and approximation algorithms from graph theory.

Friday, June 11, 1pm, 2405 Siebel Center

"The Registration Problem"

Dr. Fred Rothganger, Sandia National Laboratories

Registration is the general problem of mapping one data set so that it aligns with another. This problem occurs in various forms in many useful domains: navigation, object recognition, 3D reconstruction, and medical imaging, to name a few. This talk is a broad overview of the
registration problem, some typical ways it is formulated, and some common solutions. Along the way, it presents examples of relevant work going on in industry and at Sandia National Labs.

Fred Rothganger earned his PhD from UIUC in 2004, specializing in Artificial Intelligence and Computer Vision. Since then, he has been at Sandia National Laboratories, doing research and development in computer vision and cognitive modeling.

Thursday, June 17, 1:00pm, 2405 Siebel Center

“Data Analytics for Homeland Security”

Dr. Joseph Kielman, US Department of Homeland Security

Thursday, June 17, 2:30pm, 2405 Siebel Center

“Learning by Reading: From Information Extraction to Machine Reading”

Dr. Eduard Hovy, Information Sciences Institute / University of Southern California

Creating computer systems that educate themselves by reading text was one of the original dreams of Artificial Intelligence. Researchers in Natural Language Processing (NLP) have made initial steps in this direction, especially with Information Extraction and Text Mining, which derive information from large sets of data. Can one, however, build a system that learns by reading just one, or a small number, of texts about a given topic?

Starting in 2002, three research groups in an experiment called Project Halo manually converted the information in one chapter of a high school chemistry textbook into knowledge representation statements, and then had a knowledge representation system take the US high school standardized (AP) exam. Surprisingly, all three systems passed, albeit not very well. Could one do the same, automatically? In late 2005, DARPA funded several small pilot projects in NLP, Knowledge Representation and Reasoning (KR&R), and Cognitive Science to take up this challenge, which grew into Project Möbius, a collaboration of SRI, USC/ISI, University of Texas Austin, Boeing, and BBN Inc. The Möbius prototype learning-by-reading system read paragraph-length Wikipedia-level texts about the human heart and about engines, built up enough knowledge to apply inferences, to produce its own further reading requests, and to answer unseen questions. Results were encouraging. In 2009, DARPA funded a new 5-year program called Machine Reading, which funds three large teams that include many of the top NLP and KR&R research scientists in the USA.

This talk describes the Machine Reading program and provides details about one of the three teams, RACR, which is led by IBM's IE/QA team, and includes researchers at USC/ISI, University of Texas Austin, CMU, and the University of Utah. The system contains several reading engines that are being composed into a single large framework, with access to a cluster of several thousand computers for large-scale experiments. The reading engines include traditional Information Extraction engines, parsers, converters to various logical form
representations, abstract semantic models of world knowledge, and various kinds of abductive and other reasoning engines. I will focus on the use of large repositories of background knowledge and their various uses to support reading and inference, and describe the experiments currently being done.

Wednesday, June 23, 1pm, 2405 Siebel Center

“From Foundations to Applications of Visual Analytics”

Dr. Ross Maciejewski, Purdue University

For centuries, we have been improving our ability to collect and generate data and this is increasing at an unprecedented rate. Unfortunately, the capability to analyze and generate information and knowledge from this data has not kept pace. Over the past thirty years, the fields of visualization and information visualization have developed to help solve this problem. These active areas of research have led to the development of helpful tools for decision makers, businessman, scientists, and engineers. However, to gain insight, reason, and make decisions from these massive, disparate, uncertain, and time-evolving datasets requires more than just visualization of data. This has led to the global development of an emerging research area that goes beyond visualization: visual analytics. The science of visual analytics focuses on improving analysts abilities to understand complex problems by aiding them in gathering insight, detecting the expected and discovering the unexpected within their data. In this talk, I will provide an introduction to the fundamentals of visual analytics, describing the core components. This introduction will then shift focus to domain specific applications of visual analytics in syndromic surveillance, criminal incidence analysis, pandemic monitoring and preparedness, and more traditional scientific visualization domain problems.

Wednesday, June 23, 2:30pm, 2405 Siebel Center

“Technical and organizational challenges and solutions in Yahoo!’s massive data warehousing environment”

Yahoo’s Data Quality Team

Data is Yahoo!’s most strategic asset, and high-quality data is a key to accurate insights and monetization. This presentation describes the approach to ensure high data quality by applying recognized industry practices with a customized approach in a massive data environment. After significant data quality wins in Yahoo! data systems, three success factors are important in addressing the next level DQ issues: (1) The methodology that builds in proactive and reactive capabilities into products up-front and includes end-to-end data focus resulting in system improvements and fast issue resolution, (2) the organization of the DQ program that uses a central and embedded-in-the-businesses model with a strong focus on customer engagement, and (3) solutions for technical challenges in the internet domain in Yahoo!’s massive data environment including statistical monitoring and alerting, abuse and robot traffic detection, and latency vs. accuracy.
Friday, June 25, 1pm, 2405 Siebel Center (NOTE ROOM CHANGE)

“Relationship Extraction for Event Summarization”

Dr. William Hsu, Kansas State University

This talk will survey current open problems and approaches used in the domain of relationship extraction for summarization of events. I will first present several existing techniques for spatiotemporal event extraction and discuss the challenge of applying these techniques to domains such as epidemiology and health informatics. Information extraction tasks in this area are related to update summarization, the problem of automatically generating a brief restatement of the main points in a text when the user of the system has already read a given set of earlier articles. Next, I will survey some of the machine learning and natural language processing methodologies used in this type of event summarization. I will then describe the related task of topic detection and tracking with application to tracking current events, and discuss some relevant current work on topic modeling for this task.

Finally, I will present some results from continuing research on development of analytical tools based on this work, the emphasis of which is on performance elements such as predictive data mining, data-aware search, and question answering. I will conclude with a discussion of relevant related work in thematic mapping of relationships and opinions, focusing on link analysis approaches.

Monday, June 28, 1pm, 2405 Siebel Center (NOTE ROOM CHANGE)

“Computational Analysis of Dynamic Networks (and its applications to the social life of zebras)”

Dr. Tanya Berger-Wolf, University of Illinois-Chicago

Computation has fundamentally changed the way we study nature. Recent breakthroughs in data collection technology, such as GPS and other mobile sensors, are giving biologists access to data about wild populations that are orders of magnitude richer than any previously collected. Such data offer the promise of answering some of the big ecological questions about animal populations. The data are not unique to animal domain but is now prevalent in human interactions: emails, blogs, and online social networks. Unfortunately, our ability to analyze these data lags substantially behind our ability to collect it. In particular, interactions among individuals are often modeled as social networks where nodes represent individuals and a link exists if the corresponding individuals have interacted during the observation period. The model is essentially static in that the interactions are aggregated over time and all information about the time and ordering of social interactions is discarded.

We have extended computational methods for social network analysis to explicitly address the dynamic nature of interactions among individuals. We have developed techniques for identifying persistent communities, influential individuals, and extracting patterns of interactions in dynamic
social networks. We will present our approach and demonstrate its applicability by analyzing interactions among zebra populations.

Dr. Tanya Berger-Wolf is an assistant professor at the Department of Computer Science at the University of Illinois at Chicago where she heads the Computational Population Biology Lab. Her research interests are in applications of combinatorial optimization analysis and algorithm design techniques to problems in population biology of plants, animals, and humans, from genetics to social interactions.

Dr. Berger-Wolf has received her B.Sc. in Computer Science and Mathematics from Hebrew University (Jerusalem, Israel) and her Ph.D. in Computer Science from University of Illinois at Urbana-Champaign in 2002. She has spent two years as a postdoctoral fellow at the University of New Mexico working in computational phylogenetics and a year at the Center for Discrete Mathematics and Theoretical Computer Science (DIMACS) doing research in computational epidemiology. She has received numerous awards for her research and mentoring, including the US National Science Foundation CAREER Award in 2008 and the UIC Mentor of the Year Award in 2009.