Celebrating AAAI's 25th Anniversary
The Twentieth National Conference on Artificial Intelligence
July 9-13, 2005, Pittsburgh, Pennsylvania

AAAI's Tutorial Forum provides an opportunity for junior and senior researchers to spend two days each year freely exploring exciting advances in disciplines outside their normal focus. We believe this type of forum is essential for the cross fertilization, cohesiveness, and vitality of the AI field. We all have a lot to learn from each other; the Tutorial Forum promotes the continuing education of each member of the AAAI.

The sixteen AAAI-05 Tutorials are as follows:

Advances in Word Sense Disambiguation (SA1)
Bioinformatics and Machine Learning — An Introduction and Our Success Stories (UP1)
Building Agents in SOAR (SA2)
Downsizing Data for High Performance in Learning — Introduction to Feature Selection Methods (SA3)
Empirical Methods for Artificial Intelligence (UA1)
Exploiting Structure and Randomization for Large-Scale Constraint Reasoning (SP3)
Intelligent User Interfaces: An Introduction (UP2)
Market Clearing Algorithms (SP1)
Multiagent Planning: A Survey of Research and Applications (SA4)
Pyro: A Tool for Teaching Robotics and AI (SP2)
Robotics for Beginners; Using Robot Kits to Teach Agents and AI (UA2)
The Semantic Web (UA3)
Sensor Networks: New Challenges and Opportunities for AI (UA4)
Systematic Bounding Techniques for Combinatorial Optimization (UP3)
Text Analytics: Theory and Practice (UP4)
Where Do Heuristics Come From? (Using Abstraction to Speed Up Search) (UA5)

Advances in Word Sense Disambiguation (SA1)

Rada Mihalcea and Ted Pedersen
Saturday, July 9, 9:00am – 1:00pm

Word sense disambiguation is the problem of identifying the intended meaning (or sense) of a word, based on the context in which it occurs. This is a central problem in natural language processing, and improved approaches have the potential to advance the state of the art in natural language understanding, machine translation, information retrieval, and many other language related problems. This tutorial will introduce the full range of techniques that have been applied to this problem. These include knowledge-intensive methods that take advantage of dictionaries and other manually crafted resources, supervised techniques that learn classifiers from training examples, minimally supervised approaches that bootstrap off small amounts of labeled data, and unsupervised approaches that identify word senses in raw unannotated text. In addition, the tutorial will provide an overview of resources that are available to those who might wish to conduct research in this area, or incorporate word sense disambiguation techniques in their existing systems. The tutorial does not assume any familiarity with natural language processing on the part of attendees.

Rada Mihalcea is an assistant professor of computer science at University of North Texas. She is the president of the ACL special interest group on the lexicon (SIGLEX), and was one of the coordinators of the Senseval-3 word sense disambiguation evaluation exercise.

Ted Pedersen is an associate professor of computer science at the University of Minnesota, Duluth. He is the recipient of an NSF CAREER award, and does research in supervised, unsupervised, and knowledge intensive word sense disambiguation.

Bioinformatics and Machine Learning—
An Introduction and Our Success Stories (UP1)

Jinyan Li and Limsoon Wong
Sunday, July 10, 2:00 –6:00pm

In this tutorial we will first talk about some basic knowledge of biology, and give an introduction to bioinformatics. Then, we present an overview to prediction and data mining methods that are often used in bioinformatics studies. These methods include decision trees, committee of decision trees by bagging, boosting, random forest, randomization trees, and a recent approach of CS4; and also methods for mining closed patterns, key patterns, and emerging patterns. Then we talk about five case studies in depth, including a subtype classification of childhood leukemia, an outcome prediction using extreme patient samples, a rule-based method for insilico cancer diagnosis, a binding motif discovery from protein interaction data by fixpoint theorems, and a recognition study of gene regulation sites from genomic DNA sequences.

Limsoon Wong is the deputy executive director (research) at the Institute for Infocomm Research, Singapore. He is concurrently a senior scientist at the Institute of Molecular and Cell Biology, Singapore; and an adjunct professor of computer science at the National University of Singapore. He is currently working mostly on knowledge discovery technologies and is especially interested in their application to biomedicine. Prior to that, he conducted significant research in database query language theory and finite model theory, as well as significant development work in broad-scale data integration systems. He has authored or coauthored nearly 100 papers in refereed journals and edited volumes in these areas, and has served on the program committees of many international research conferences in these areas. He is a managing editor of the *Bioinformatics and Computational Biology* journal and an advisory editor of *Drug Discovery Today*—the most heavily cited journal in the field of drug discovery. He received his BSc(Eng) in 1988 from Imperial College London and his Ph.D. in 1994 from University of Pennsylvania, both in computing.

Jinyan Li is deputy head of the Decision Systems Laboratory at Institute for Infocomm Research, Singapore. His research interests include data mining, machine learning, and bioinformatics. He is a member of the winning team for the Gold Award at the 2003 Asia Innovation Awards Competition for the significant contribution to a single test platform for accurate prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling.

Building Agents in SOAR (SA2)

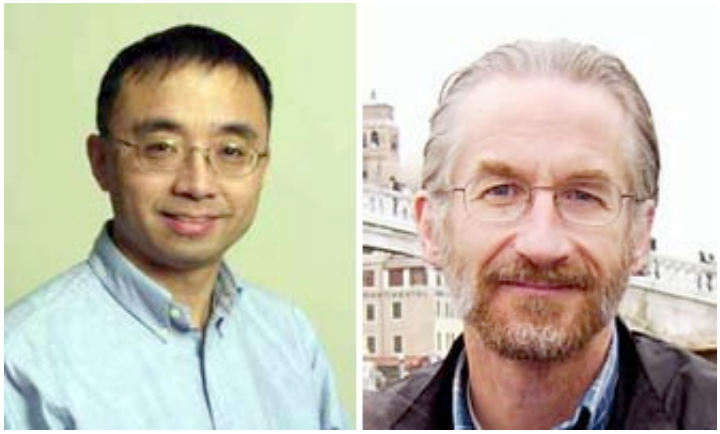
John E. Laird
Saturday, July 9, 9:00 am –6:00 pm (All Day)

This full-day tutorial will provide participants an understanding of the details of Soar so that they can create simple Soar programs. This will be a full-day tutorial, starting with a history of Soar, a comparison of it to related architectures, and an analysis of its strengths and weaknesses. The remainder of the tutorial will be hands-on. In the morning, participants will learn to run, modify, and debug small demonstration programs that illustrate the various parts of Soar's structure, including it new reinforcement learning component. They will also be introduced to Soar's editing, debugging, and runtime tools. In the afternoon, we will work on simple agents that interact with a dynamic simulated environment. The students will build their own complete agents that navigate and compete in a simple maze world. The final portion of the tutorial, the students will be exposed to the implementation of more advanced topics, including planning, metareasoning, and different forms of learning. For this part, the students will run existing programs, make only small changes, but will not do extensive programming.

Prerequisite knowledge: We expect participants to have general programming experience and a basic understanding of AI. No prior knowledge of Soar or rule-based systems required. Students will need to bring their own computer to participate in the tutorial. All required software will be provided for download prior to the tutorial.

John E. Laird is a professor of electrical engineering and computer science at the University of Michigan. He received his Ph.D. in computer science from Carnegie Mellon University in 1983. He is one of the original developers of Soar and leads its continued development and evolution. He is a Fellow of AAAI.

Downsizing Data for High Performance in Learning — Introduction to Feature Selection Methods (SA3)



Huan Liu and Robert Stine
Saturday, July 9, 9:00 am –1:00 pm

Feature selection refers to the problem of selecting attributes or variables to be used to construct models that describe data for various purposes. These purposes range from reducing dimensionality to improving the predictive accuracy of algorithms to removing irrelevant and/or redundant features. The challenge of feature selection is most acute when confronted by "wide" data sets, data sets that support a vast number of models. Data from microarray experiments, transaction logs, and Web sites pose a pressing challenge to methods that aim to turn data into business intelligence or useful knowledge. This tutorial introduces the basic concepts, heuristics, and techniques of feature selection, surveys the current state of rapidly moving research in feature selection, reviews the state of the art algorithms and their applications (customer relationship management, credit modeling, fraud detection, gene expression microarray analysis, text analysis, and Web mining), and presents research challenges and anticipates future work. Important issues such as conceptual models of feature selection, search, validation, performance criteria, scalability, and overfitting will be discussed.

The audience is expected to have minimum knowledge in data structures and algorithms, search, and statistics. Ideally, the audience has some experience related to or the need for dealing with high-dimensional data.

Huan Liu is on the faculty of Arizona State University and received his Ph.D. from the University of Southern California. His research focuses on developing methods for feature selection, data reduction, machine learning, and data mining applications. He has published many papers and books on these topics in AI and machine learning literature.

Robert Stine is a professor of statistics at the Wharton School at the University of Pennsylvania. His research develops methods for selecting features from large databases used in predictive modeling. He then applies these methods in business applications including marketing of new products and assessing retail credit risk. His research has appeared in numerous journals and books.

Empirical Methods for Artificial Intelligence (UA1)

Paul Cohen
Sunday, July 10, 9:00 am –1:00 pm

The purpose of this tutorial is to introduce attendees to empirical tools including experiment and metrics design, exploratory data analysis, statistical hypothesis testing, and modeling. Through case studies and other examples, this tutorial introduces a variety of evaluation methods, including methods for visualizing, summarizing and detecting clues in data; for designing controlled experiments; and for testing hypotheses, both with classical tests and with new, computer-intensive bootstrap and randomization methods. The case studies are organized around ten themes, each a "mantra" that I find helpful when conducting evaluations (e.g., "follow the variance," and "evaluation begins with claims.").

The tutorial does not require any prior knowledge of the subject, and specifically will not require background in statistics. The tutorial will therefore is suitable for the general AI audience, both academic and industrial.

Paul R. Cohen is deputy director of the Intelligent Systems Division at USC's Information Sciences Institute and Research professor of computer science at USC. His Ph.D. is from Stanford University in computer science and psychology, in 1983; and his M.S. and B.A. degrees in psychology are from University of California, Los Angeles and the University of California, San Diego, respectively. He is author of *Empirical Methods for Artificial Intelligence* (MIT Press, 1995).

Cohen advises several government agencies on the design of evaluations for programs, and has published widely in methodology and in several areas of AI.

Exploiting Structure and Randomization for Large-Scale Constraint Reasoning (SP3)



Carla Gomes, Brian Williams, and Martin Sachenbacher

Saturday, July 9, 2:00 –6:00 pm

In recent years we have made tremendous strides in the design of search engines for reasoning tasks such as constraint satisfaction problems (CSP), Boolean satisfiability problems (SAT), finite constraint optimization problems (COP), diagnosis, planning, execution and bounded model-checking. For example, in areas, such as planning and bounded model-checking, we are now able to solve large CSP's with up to a million variables and five million constraints. Likewise, we are robustly controlling autonomous space probes, cooperating air vehicles, and biped robots, by solving large COP's within an autonomous system's reactive control loop. In this tutorial we will focus on progress that has been driven by exploiting special problem structure, randomization and approximation techniques, constraint learning, problem decomposition methods, and incremental methods. Special structure can be exploited using, for example, network flow algorithms, linear programming-based relaxations, streamlining constraints, and symbolic encodings. Randomization techniques exploit heavy-tailed phenomena in combinatorial search. Constraint learning methods uncover new structure during problem solving. Incremental methods achieve efficiency by reasoning about the effects of small problem perturbations. Problem decomposition methods, such as hyper tree decomposition, partition a problem into more manageable, lower dimensional sub problems. Finally, symbolic encodings reason efficiently about sets of solutions, by exploiting a compact state space encoding.

The tutorial will include real-world applications in the area of scheduling, planning, diagnosis and autonomous systems control.

Carla Gomes received her Ph.D. in computer science, in the area of artificial intelligence and operations research, from the University of Edinburgh. Her research has covered several areas in artificial intelligence and computer science, including planning and scheduling, integration of CSP and OR techniques for solving combinatorial problems, and algorithm portfolios. Carla Gomes is an associate professor of computing and information science and applied economics and management at Cornell University. She is also the director of the Intelligent Information Institute at Cornell.

Brian Williams received his S.B., S.M and Ph.D. from MIT in computer science and artificial intelligence in 1989. He pioneered multiple fault, model-based diagnosis in the 1980s through the GDE and Sherlock systems at the Xerox Palo Alto Research Center, and model-based autonomy in the 1990s through the Livingstone model-based health management and the Burton model-based execution systems. At the NASA Ames Research Center from 1994 to 1999 he formed the Autonomous Systems Area, and coinvented the Remote Agent model-based autonomous control system, which received a NASA Space Act Award in 1999. He was a member of the NASA Deep Space One probe flight team, which used remote agent to create the first fully autonomous, self-repairing explorer, demonstrated in flight in 1999. He has won two best paper prizes for his research in qualitative algebras and fast propositional inference. He was a member of the Tom Young Blue Ribbon Team in 2000, assessing future Mars missions in light of the Mars Climate Orbiter and Polar Lander incidents. He has served as guest editor of the *Artificial Intelligence Journal* and has been on the editorial boards of the *Journal of Artificial Intelligence Research,* and The MIT Press. He is currently a member of the Advisory Council of the NASA Jet Propulsion Laboratory at Caltech.

Martin Sachenbacher is a postdoctoral associate at the MIT Computer Science and Artificial Intelligence Laboratory. He received his Ph.D. from Technische Universitaet Muenchen, Germany, in 2001 and worked in several projects involving major car companies. He pursues research in model-based programming, constraint optimization, diagnosis, and qualitative modeling.

Intelligent User Interfaces: An Introduction (UP2)

Mark Maybury
Sunday, July 10, 2:00 –6:00 pm

Intelligent user interfaces (IUI) aim to improve human-machine interaction by representing, reasoning, and intelligently acting on models of the user, domain, task, discourse, and media (e.g., graphics, natural language, gesture). IUIs are multifaceted, in purpose and nature, and include capabilities for multimedia input analysis, multimedia presentation generation, and the use of user, discourse and task models to personalize and enhance interaction. Some IUIs support asynchronous, ambiguous, and inexact input through analysis of multimodal input. Others include animated computer agents that express system and discourse status via facial displays, that tailor explanations to particular contexts, or that manage dialogues between human and machine.

Potential IUI benefits include: (1) More efficient interaction—enabling more rapid task completion with less work. (2) More effective interaction—doing the right thing at the right time, tailoring the content and form of the interaction to the context of the user, task, dialogue and (3) More natural interaction—supporting human-like spoken, written, and gestural interaction.

Drawing upon material from *Readings in Intelligent User Interfaces* (Maybury and Wahlster, 1998), this tutorial will define terms, outline the history, describe key subfields, and exemplify and demonstrate intelligent user interfaces in action.

Mark Maybury is executive director of MITRE's Information Technology Division. Mark also serves as executive director of ARDA's Northeast Regional Research Center (nrrc.mitre.org). Maybury has published over fifty technical and tutorial articles and is editor of *Intelligent Multimedia Interfaces* (AAAI/MIT Press, 1993), *Intelligent Multimedia Information Retrieval* (AAAI/MIT 1993), *Intelligent* (AAAI/MIT Press, 1997), *New Directions in Question Answering* (AAAI/MIT Press, 2005), coeditor of *Readings on Intelligent User Interfaces* (Morgan Kaufmann Press, 1998), *Advances in Text Summarization* (MIT Press, 1999) and *Advances in Knowledge Management* (MIT Press, 1999) and coauthor of *Information Storage and Retrieval* (Kluwer Academic, 2000). Maybury received his B.A. in

mathematics from the College of the Holy Cross, an M. Phil. in computer speech and language processing from Cambridge University, England, an M.B.A. from Rensselaer Polytechnic Institute and a Ph.D. in artificial intelligence from Cambridge University.

Market Clearing Algorithms (SP1)

Tuomas Sandholm
Saturday, July 9, 2:00 –6:00 pm

Markets are important mechanisms for allocating goods, services, tasks, and resources among multiple agents, be they human or software. The market clearing problem is that of deciding how to allocate the items among the agents. The last five years have witnessed a leap of improvement in market clearing algorithms both for traditional market designs and entirely new market designs enabled by advanced clearing technology. This tutorial covers the computational implications of different market designs and presents algorithms for clearing markets optimally and approximately. Auctions (one seller, multiple buyers), reverse auctions (one buyer, multiple sellers), and exchanges (multiple buyers and multiple sellers) are covered. Both theoretical and experimental results are presented. Multiitem and multiunit markets will be a key focus. Computational implications of different classes of side constraints will be presented. Bid types covered include price-quantity bids, different shapes of supply/demand curves, and package bids. Methods for selective incremental preference elicitation for combinatorial markets are presented, with which the market can be cleared optimally using only a small portion of the agents' preferences as input.

A basic understanding of algorithms, search, and NP-completeness is necessary. No background on markets is assumed.

Tuomas Sandholm is an associate professor in the Computer Science Department at Carnegie Mellon University. He received his Ph.D. and M.S. degrees in computer science from the University of Massachusetts at Amherst in 1996 and 1994. He earned an M.S. (B.S. included) with distinction in industrial engineering and management science from the Helsinki University of Technology, Finland, in 1991. He has 15 years of experience building electronic marketplaces, and has fielded over 150 large-scale combinatorial auctions. He has published over 200 papers, and received the inaugural ACM Autonomous Agents Research Award, the NSF Career award, the Sloan Fellowship, and the IJCAI Computers and Thought award.

Multiagent Planning: A Survey of Research and Applications (SA4)

Bradley J. Clement and Keith Decker
Saturday, July 9, 9:00 am –1:00 pm

Multiagent planning is concerned with planning by (and for) multiple agents. It can involve agents planning for a common goal, an agent coordinating the plans or planning of others, or agents refining their own plans while negotiating over tasks or resources. Distributed continual planning addresses these problems when further complicated with interleaved execution. More than ever industry, space, and the military are seeking systems that can solve these problems.

This tutorial will describe variations of the multiagent planning problem, discuss issues in the applicability and design of multiagent planning systems, and describe some real-world multiagent planning problems. We will also review the history of research contributions to this subfield and describe frameworks and systems such as Distributed NOAH, GPGP, DSIPE, and SHAC. In addition, we will describe open research issues in multiagent planning and its overlap and relation to other fields, such as market-based AI and game theory.

Brad Clement is a senior member of the Artificial Intelligence Group at the Jet Propulsion Laboratory in Pasadena, CA, where he is developing methods for coordinating planning and scheduling for single and multiple spacecraft/missions. He received a Ph.D. in computer science and engineering from the University of Michigan, Ann Arbor.

Keith Decker is an associate professor of computer and information sciences at the University of Delaware, where he continues work on the DECAF agent toolkit and extending TAEM.S. and GPGP approaches to coordinated planning and scheduling. He received his Ph.D. in computer science from the University of Massachusetts, Amherst.

Pyro: A Tool for Teaching Robotics and AI (SP2)

Holly Yanco and Doug Blank
Saturday, July 9, 2:00 –6:00 pm

Pyro (Python robotics) is an open-source software system that abstracts away the details of particular robots and allows users to explore complex control methods at a high level. Until now, it has been necessary to learn very different and specific control languages for different mobile robots, particularly those manufactured by different companies. Now, a single language can be used to program many different robots, allowing code to be shared across platforms.

Currently, the robots supported include the Pioneer family, the Khepera family, and the Sony AIBO. Supported simulators include Player/Stage, Gazebo, the Aria simulator, the Khepera simulator and the RoboCup simulator.

This hands-on tutorial will introduce the abstractions provided by Pyro, then will show participants how to get started with running Pyro and writing robot brains. We will present an overview of the modules of different control paradigms and AI methods. Each participant will be given a bootable CD (on an Intel architecture) containing the Pyro software to take home.

The tutorial is intended for people planning to teach a course in robotics or artificial intelligence and for other robotics and AI researchers who want to learn a new tool for exploring robotics and AI topics.

Holly Yanco and Douglas Blank are coprincipal investigators on “Beyond LEGOs: Hardware, Software and Curriculum for the Next Generation Robotics Laboratory,” NSF DUE-0231363, which funded the development of Pyro.

Holly Yanco graduated from MIT with a Ph.D. in computer science in 2000. She is an assistant professor in the Computer Science Department at the University of Massachusetts at Lowell. Her research interests include human-robot interaction, adjustable autonomy, and assistive technology. She has received teaching awards from UMass Lowell and MIT.

Douglas Blank graduated from Indiana University with a joint Ph.D. in cognitive science and computer science in 1997. He is an assistant professor of computer science at Bryn Mawr College. His research interests include emergent systems, connectionist models of analogy-making, pedagogical tools for teaching robotics, gender issues in computer science, and developmental models in robotics.

Robotics for Beginners; Using Robot Kits to Teach Agents and AI (UA2)

Elizabeth Sklar and Simon Parsons
Sunday, July 10, 9:00 am –6:00 pm (All Day)

Robots have always been prototypical agents; they are situated in the physical world and have all the classic problems of noisy sensors and imprecise actuators. Through kits like the LEGO Mindstorms, agents researchers have access to robots much more easily than before, both for teaching and research. The kits are being used in introductory robotics classes, even in departments where there is no robot research lab. Some faculty members are bringing the kits into classes that are not specifically about robots, but as a means for giving students a hands-on platform with which to experiment with the some of the abstract concepts they are learning.

The purpose of this tutorial is to show how the LEGO Mindstorms kit can be used as a means for exploring AI and agents topics, such as individual agent control architectures, swarmlike group behaviors and agent communication mechanisms. The full-day session will begin with an overview of the basic concepts relating to intelligent agents and autonomous robotics, followed by instruction on the use of the hardware and software, and finishing with simple exercises to give attendees hands-on experience with the robotics kits (provided by the presenters).

The tutorial is intended for any delegates who have some basic knowledge of AI and the C programming language. No prior experience with LEGO or LEGO robotics kits is required.

Elizabeth Sklar is an assistant professor of computer science at Columbia University. Her research studies interactive learning systems, educational robotics and multiagent simulation. She received her Ph.D. from Brandeis University (2000) and received the RoboCup Scientific Challenge Award (2002). She is a RoboCup Trustee and has served on committees for RoboCup, AAAI, AAMAS.

Simon Parsons is an associate professor of computer and information science at Brooklyn College, City University of New York. He received his Ph.D. from University of London (1993) and was awarded the 1998 IEE Younger Engineers Achievement Medal. He has published over 100 papers on decision making in intelligent systems and is the editor of the *Knowledge Engineering Review.*

The Semantic Web (UA3)

Deborah L. McGuinness and Mike Dean
Sunday, July 10, 9:00 am –1:00 pm

The semantic web can be viewed as the next generation web—one that enables smarter, more useful, customizable, and interoperable applications. The differentiator between the web and the semantic web is an encoding of the meaning of terms that will be used in applications. This tutorial will introduce the semantic web with a brief historical perspective, provide an introduction to the primary languages for encoding meaning, and introduce ontologies and show how they can and are being used to power to the semantic web. We will also describe and demonstrate a number of the tools available today to support the Semantic Web. Tools will address the topics of ontology creation, reading and writing content, validation, reasoning, explanation, storage, query, and database integration. We will include a tour of the open source software repository SemWebCentral.org and include a description of current semantic web research themes.

Deborah L. McGuinness is codirector and senior research scientist at the Knowledge Systems Division (KSL) of the Artificial Intelligence Laboratory at Stanford University. She has 25 years of experience in knowledge representation and reasoning environments for ontology creation and maintenance. She has built and deployed numerous ontology environments and ontology applications. She is coeditor of the World Wide Web Consortium Recommendation Ontology Web Language (OWL) and has worked in the semantic web since its inception. She helped start the web ontology working group out of work as a co-author of the DARPA Agent Markup Language program's DAML language. She is a leader in ontology-based tools and applications. In addition to leading research efforts in semantic web, explanation, proof, trust, and representation and reasoning languages, McGuinness consults with companies interested in deploying semantic-enabled applications. She helped VerticalNet design and build its Ontobuilder/Ontoserver ontology evolution environment, helped GM build its ontology-enhanced search component for its Variation Reduction effort that won an IAAI award in 2004, and helped CISCO with its Meta Data effort.

Mike Dean is a principal engineer at BBN Technologies and principal investigator for the DARPA Agent Markup Language (DAML) Integration and Transition effort. He chairs the Joint US/EU ad hoc Agent Markup Language Committee and was a member of the W3C RDF Core and Web Ontology Working Groups. He is the author or coauthor of a number of semantic web tools, including PalmDAML, DAML DB, and SweetRules. He currently leads the development of several semantic web pilot applications for government organizations.

Sensor Networks: New Challenges and Opportunities for AI (UA4)

Carlos Guestrin
Sunday, July 10, 9:00 am –1:00 pm

Sensor networks consist of a collection of small low-cost devices that can sense and actuate in their environment, and communicate with each other through a wireless network. Recent advances in hardware and low-level software have made it possible for several real-world deployments to collect scientific and engineering data in unstructured environments. The long-term goal of sensor networks is to design systems that address significantly more complex applications, including fault detection in large-structures, such as bridges, building automation, smart environments, and management of disaster rescue operations. Most of these future applications require the solution of complex data analysis, learning, and decision making tasks. Recent advancements in probabilistic AI algorithms and representations seek to solve such complex tasks, but sensor networks pose unique new challenges preventing such direct applications of existing methods: Resources, such as computation, communication and power, are usually scarce in low-cost wireless sensor networks. Furthermore, communication is usually unreliable, and nodes are prone to failures. In this tutorial, we will discuss these challenges, current solutions, and the opportunities for fundamentally new AI algorithms that address the real issues arising in sensor networks. The content is accessible to attendees with working knowledge of probabilistic methods in AI or in robotics.

Carlos Guestrin is an assistant professor at Carnegie Mellon University. Previously, he was a senior researcher at the Intel Berkeley Lablet. Carlos received his Ph.D. in computer science from Stanford University. His current research interests span planning, reasoning and learning in uncertain dynamic environments, focusing on applications in sensor networks.

Systematic Bounding Techniques for Combinatorial Optimization (UP3)


Sharlee Climer and Weixiong Zhang
Sunday, July 10, 2:00 –6:00 pm

Combinatorial optimization problems are ubiquitous in scientific research, engineering, and even our daily lives. A major research focus in developing combinatorial search algorithms has been on the attainment of efficient methods for deriving tight lower and upper bounds. These bounds restrict the search space of combinatorial optimization problems and facilitate the computation of what might otherwise be intractable problems.

In this tutorial, we survey the history of the use of bounds in both AI and operations research. While research has been extensive in both domains, until very recently it has been too narrowly focused and has overlooked great opportunities to exploit bounds.

In the past, the focus has been on the relaxations of constraints. We discuss methods for deriving bounds by tightening constraints, adding or deleting decision variables, and modifying the objective function. Then a formalization of the use of bounds as a two-step procedure is covered. Finally, we discuss recent developments demonstrating how the use of this framework is conducive for eliciting methods that go beyond search-tree pruning.

This tutorial is aimed at the general AI audience. Familiarity with some basic concepts of combinatorial optimization is desirable but not essential.

Weixiong Zhang is an associate professor at Washington University in St. Louis. He has more than ten years of research experience in the areas of AI (such as heuristic search, phase transitions, and multiagent systems), combinatorial optimization (for example, the traveling salesman problem) and computational biology (sequence alignment, motif finding, and RNA folding).

Sharlee Climer is a doctoral candidate at Washington University in St. Louis, anticipating graduation in May 2005. She holds degrees in computer science, physics, structural engineering, and mathematics. Her dissertation and several publications are based on the exploitation of bounds. Climer is an NDSEG Fellow and an Olin Fellow.

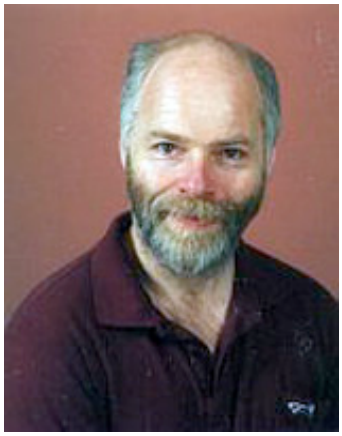
Text Analytics: Theory and Practice (UP4)

Ronen Feldman
Sunday, July 10, 2:00 –6:00 pm

Text Analytics is a new and exciting research area that tries to solve the information overload problem by using techniques from data mining, machine learning, NLP, IR and knowledge management. Text analytics involves the preprocessing of document collections (text categorization, information extraction, term extraction), the storage of the intermediate representations, the techniques to analyze these intermediate representations (distribution analysis, clustering, trend analysis, association rules etc) and visualization of the results. In this tutorial we will present the general theory of Text analytics and will demonstrate several systems that use these principles to enable interactive exploration of large textual collections. We will present a general architecture for text analytics and will outline the algorithms and data structures behind the systems. Special emphasis will be given to efficient algorithms for very large document collections, tools for visualizing such document collections, the use of intelligent agents to perform text analytics on the internet, and the use information extraction to better capture the major themes of the documents. The Tutorial will cover the state of the art in this rapidly growing area of research. Several real world applications of text analytics will be presented.

Ronen Feldman is a senior lecturer at the Mathematics and Computer Science Department of Bar-Ilan University in Israel, and the Director of the Data Mining Laboratory. He received his B.Sc. in mathematics, physics, and computer science from the Hebrew University, and his Ph.D. in computer science from Cornell University in New York. He is the founder, president, and chief scientist of ClearForest Corporation.

Where Do Heuristics Come From?
(Using Abstraction to Speed Up Search) (UA5)



Robert Holte
Sunday, July 10, 9:00 am –1:00 pm

This tutorial is aimed at AI researchers and students interested in methods for creating heuristics to be used by algorithms such as A* and IDA*. The tutorial will focus primarily on pattern databases, a general-purpose technique for defining heuristics that has revolutionized heuristic search, leading to breakthroughs such as the first optimal solutions for random instances of Rubik's Cube. Related techniques, such as hierarchical search, will also be discussed. Pattern databases store a heuristic in the form of a lookup table, with the heuristic itself being defined as the true distance from a state to the goal in an abstract (simplified) version of the state space. Using abstraction to define a heuristic has two key advantages: (1) very little domain-specific knowledge is needed to define accurate heuristics, which is particularly important for novel or one-time problems; and (2) numerous different high-quality heuristics can be created for a give state space.

The tutorial will start with basic concepts and proceed through to the most recent advances. It assumes no background other than a familiarity with A* or IDA*.

Robert Holte is a former editor-in-chief of *Machine Learning* and current director of the Alberta Ingenuity Centre for Machine Learning. His current research investigates cost-sensitive learning, learning in poker and commercial computer games, and the use of abstraction to speed up single-agent search.