IEEE Intelligent Systems
www.computer.org/intelligent

AI's 10 to Watch

Philipp Cimiano, Dmitri Dolgov, Anat Levin, Brian Milch, Peter Mika, Louis-Philippe Morency, Boris Motik, Jennifer Neville, Erik Sudderth, and Luis Von Ahn

Vol. 23, No. 3
May/June 2008

AI's 10 to Watch

The field of computer science is in a struggle to define its core in a ubiquitous-computing world where pervasive computer use and end-user programming give us cause to question some basic assumptions of our paradigms. In AI, however, our intellectual challenges remain largely unchanged. The quest to understand intelligence, whether by cognitive emulation or engineering fiat, remains one of the true frontiers available to modern science. Despite increasingly powerful computers, robots' growing capabilities, and the development of new and better methods for combining information, the quest for intelligence continues to elude us. The need remains constant for creative leaders who will take us into new AI paradigms or find ways to innovate in those paradigms we're already exploring.

IEEE Intelligent Systems started "AI's 10 to Watch" to acknowledge, early in their careers, the pioneers and leaders of the future. The 10 researchers we highlight here have all completed their doctoral work in the past couple of years. We cast a wide net for recommendations, asking the chairs of computer science departments, the AAAI Fellows, and all our readers for nominations. We received many nominations, all of them worthy of note. A committee comprising members of our advisory and editorial boards reviewed all the applications, and we unanimously agreed on the top 10, whom we present to you in this special section. As I think you'll agree after reading these profiles, our field's future is in excellent hands.

—*James Hendler*

1541-1672/08/$25.00 © 2008 IEEE
Published by the IEEE Computer Society

A I ' s 1 0 t o W a t c h

Philipp Cimiano

University of Karlsruhe

Philipp Cimiano is a postdoctoral researcher and research project leader at the University of Karlsruhe's Institute AIFB (Institute of Applied Informatics and Formal Description Methods). He graduated in computer science and computational linguistics from the University of Stuttgart and received his PhD in applied computer science from the Technical University of Karlsruhe. His PhD research aimed to develop algorithms for learning ontologies from text corpora and from the Web. This research ultimately led to the publication of *Ontology Learning and Population from Text: Algorithms, Evaluation and Applications* (Springer, 2006). Cimiano's research interests span a wide range of areas, including knowledge representation, the Semantic Web, and ontologies, in particular, as well as computational linguistics, text mining, and machine learning. He's an active member of the research community, serving on conference program committees and as a reviewer for journals in all the previously mentioned fields. He received a Best Presentation Award for his talk at the 2004 World Wide Web Conference. When he's not busy writing papers or teaching, he likes sports, such as cross-country skiing, hiking, or running marathons. Contact him at cimiano@aifb.uni-karlsruhe.de.

AI and Natural Language

Understanding natural language has been a traditional goal of AI since the beginnings of the field. Full understanding of natural language implies human-level capability to process language and draw appropriate conclusions. Needless to say, this task is not only very complex—if feasible at all—but also quite unbounded in terms of scope. The crucial question is, what level of understanding and complexity of drawn conclusions should we require from a machine?

My working hypothesis is that the task of "understanding" natural language becomes feasible if we clearly restrict the scope of what must be understood. To do this, we can focus on a specific domain model and specify which conclusions a system must draw for it to successfully perform a given task.

My goal is to make computers understand language to a limited but well-defined degree, required by a certain application or domain, for example. The granularity and structural complexity of what a system needs to understand is clearly determined by the domain in question, which we assume is modelled as an ontology. We can then exploit the domain's particularities to guide semantic interpretation—for example, by helping disambiguate in the case of multiple possible interpretations. Such an ontology-based approach to language processing has important applications not only in question answering, information extraction, and dialogue systems but also in the Semantic Web, which requires that today's Web content is formalized using ontologies.

Ontology-based natural-language-understanding systems must be developed in a principled way so that they easily port to other domains or applications. When trying to transfer such systems into practice, we face several important questions. How much does the system cost? How much time must we devote to customizing it for a specific purpose? Which expertise do we need to customize it? Which methodology should we follow to adapt it to a specific purpose? Are tools available to support system debugging and maintenance? In the fields of computational linguistics and AI, we've focused so far on generic and domain-independent solutions, such as parsing and disambiguating general newspaper text or techniques for generic inference. We have, however, no clear answers to the previous questions; consequently, the transfer of research results into practice has been somewhat limited.

To answer these questions, we must consider natural-language-processing technology from the usability viewpoint. This involves developing principled systems that nonexperts can easily adapt to domains or applications, following a clearly defined and empirically verified methodology. This also involves providing tools that support the different methodological steps, particularly system debugging and maintenance. Systems will have an impact in practice only if

- they're easy to adapt to different purposes and
- they show predictable behavior that can be straightforwardly monitored by nonexperts and be changed in a controlled fashion.

Working toward developing such systems is one of my personal challenges for the coming years.

AI for Autonomous Robotic Cars

The concept of driverless robotic cars has been on AI researchers' radar for several decades. With the recent advances in many areas of computing, this vision could now become a reality. Autonomous driving is an exciting research domain that offers challenges in various areas of AI and robotics, such as perception and sensor fusion, mapping and localization, modeling and learning, and prediction and decision making. My current research touches on several of these topics, with the goal of developing algorithms that let robotic vehicles sense and understand the environment and drive safely in accordance with human rules and conventions.

Recently, one of my main interests has been path planning for autonomous navigation in unknown environments. The task here is to generate trajectories in real time in response to new information about the environment obtained by the robot's onboard sensors. A main challenge in developing a practical path planner arises from the fact that the space of all vehicle controls (and thus trajectories) is continuous. So, the task of computing near-optimal trajectories that are safe and smooth and that satisfy the vehicle's kinematic constraints becomes a complex nonlinear continuous-variable optimization problem.

My colleagues from the Stanford Racing Team (http://cs.stanford.edu/group/roadrunner) and I have developed a path-planning algorithm that addresses this computational problem by combining heuristic graph search, potential fields, and numerical continuous-variable optimization. We've successfully tested this research in the Darpa Urban Challenge (www.darpa.mil/grandchallenge), in which our vehicle flawlessly performed maneuvers in free-navigation environments.

Another research direction that's of great interest to me and, in fact, a prerequisite for good decision making is the problem of sensing and understanding the environment. A main challenge in this area is the sparsity and noise in the sensor data, which lead to an inherently ambiguous problem and force the use of probabilistic inference techniques for perception and mapping. Along this direction, my colleagues and I recently have been investigating ways to infer global properties of the environment on the basis of local observations provided by the sensors. For example, many human-built environments (such as parking lots) are well structured. So, I've been working on methods for constructing a globally consistent model of that structure by extracting local geometric features from sensor data and then using probabilistic techniques to combine them into a global view.

Despite the tremendous breakthroughs that have been made in autonomous driving in recent years, many unsolved problems remain. Among those that especially interest me are reliable identification and tracking of dynamic objects, predicting their motion, and using that information when making decisions. Making progress on these tasks will require advances on several fronts, including robust perception techniques, methods for learning and modeling dynamical systems, and efficient planning and decision-making algorithms. My goal over the next several years is to continue tackling these problems, using probabilistic AI techniques.

AI's 10 to Watch

Dmitri Dolgov

Toyota Research Institute

Dmitri Dolgov is a senior research scientist in the Toyota Research Institute's AI & Robotics group and a visiting research scientist at Stanford University's AI lab. He received a BS and an MS in applied physics and mathematics from the Moscow Institute of Physics and Technology. Dolgov also received an MS and a PhD in computer science from the University of Michigan for work on algorithms for decision making in stochastic multiagent systems with limited shared resources. His current research focuses on creating autonomous robotic vehicles. Contact him at ddolgov@ai.stanford.edu.

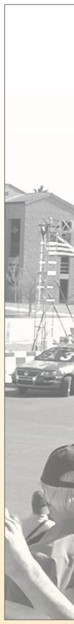


Image Statistics in Computational Photography

The digital-photography revolution dramatically simplified how we take and share pictures. Yet, digital photography sticks mostly to the rigid imaging model inherited from traditional photography.

The emerging field of *computational photography* goes one step further and exploits digital technology to enable arbitrary computation between the light array and the final image or video. Such computation can overcome the imaging hardware's limitations and enable new applications.

Computational photography opened the door for an emerging family of creative postprocessing manipulations, such as editing color and illumination, removing and pasting objects from images, or changing depth-of-field effects. Beyond postexposure manipulations, computational photography lets us insert computation into the optical design and exploration of less traditional optical setups. The rapidly developing family of computational cameras has the potential to capture additional dimensions of the signal such as depth and material reflectance properties.

Computational photography forces us to face fundamental questions about the nature of the images around us. A key issue involved with the development of stronger computational-photography tools is that images are much more than arbitrary random arrays of numbers. So, I'm interested in understanding what an image is and what makes natural images (images we see in the world around us) special.

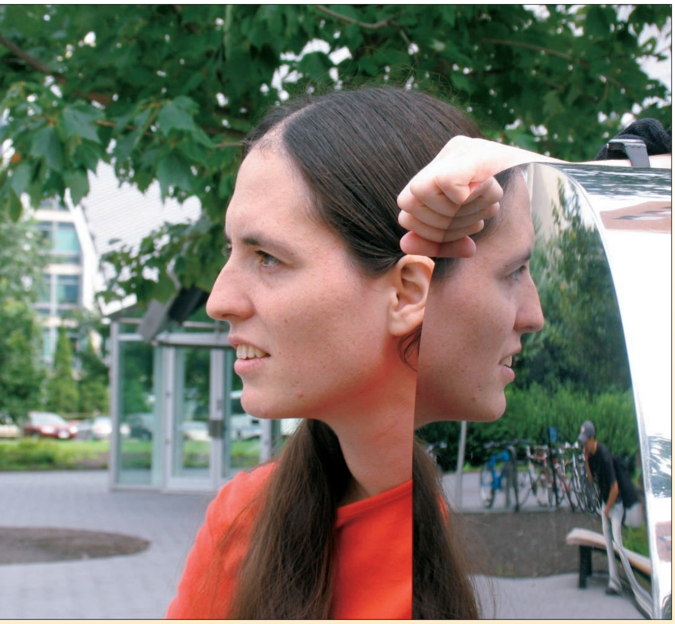
We can approach understanding images at the high level of their content, but even the low-level properties of image patterns obey unique and powerful statistical correlations. Low-level image priors are critical because they'll enable us to infer better and more natural images. Additionally, priors on the signals our camera will likely measure can help us design better cameras by predicting which dimensions of the signals are the most important to capture and how to fill the missing dimensions in the decoding process.

This understanding guided my colleagues and me in the development of several recent postexposure applications such as transparency, colorization, matting, and segmentation.

More recently, we've used image priors for a simultaneous design of both coding optics and postexposure decoding. We've demonstrated a coded-aperture camera—a new, simple change to conventional lenses that facilitates the acquisition of depth in addition to a full-resolution image from a single shot.

I believe that the principles of information budget and natural-image statistics will become more and more central to the development of future cameras.

AI's 10 to Watch



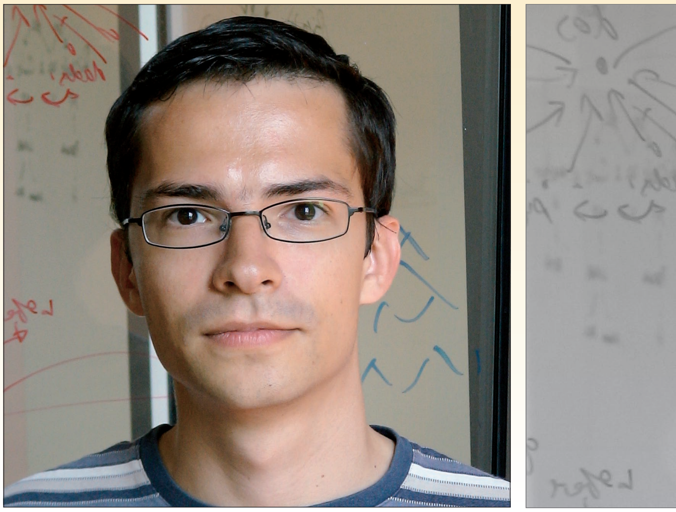
Anat Levin

MIT Computer Science and Artificial Intelligence Laboratory

Anat Levin is a postdoctoral associate at the Massachusetts Institute of Technology's Computer Science and Artificial Intelligence Laboratory. In 2009, she will join the Weizmann Institute of Science's computer science faculty. Her research interests include computer vision, computer graphics, and machine learning. In particular, she has worked on low- and mid-level vision, computational photography, and visual recognition. Levin received her BSc, MSc, and PhD in computer science from the Hebrew University of Jerusalem. Contact her at alevin@csail.mit.edu; http://people.csail.mit.edu/alevin.

Peter Mika

Yahoo! Research



Peter Mika is a researcher at Yahoo! Research in Barcelona. His interdisciplinary work in social networks and the Semantic Web earned him a Best Paper Award at the 2006 International Semantic Web Conference and a First Prize at the 2005 Semantic Web Challenge. He has been cochair of the Semantic Web Challenge since 2007. Mika is the youngest member elected to the editorial board of the *Journal of Web Semantics*. He's the author of *Social Networks and the Semantic Web* (Springer, 2007). He received his BS in computer science from Eötvös Loránd University and his MSc and PhD in computer science (cum laude) from Vrije Universiteit Amsterdam. Contact him at pmika@yahoo-inc.com.

AI's 10 to Watch

Lighting Up the Semantic Web

Being a young AI researcher comes with advantages of its own. I'm too young to have experienced the AI Winter firsthand. Yet, I've inherited from previous generations of AI researchers an incredible set of tools that let us model knowledge and many aspects of human reasoning to perform practical tasks. I can say that I'm a member of the first generation that has grown up with a computer (smuggled into communist Hungary piece-by-piece by members of my family). I was also the first in my high school to experience the Web. In addition, I was sitting in one of the first university courses ever about Tim Berners-Lee's grand vision of the Semantic Web.

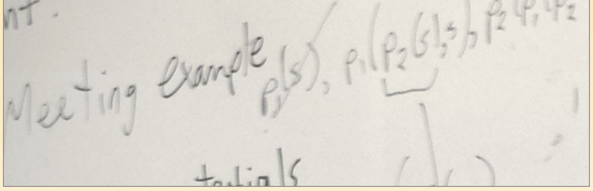
By the time I was beginning my PhD, the Semantic Web's technological foundations were already being laid down, based on the results of decades of research into knowledge representation and reasoning. As a pragmatist, however, I was most excited about what would happen next: how would we fill the Semantic Web with life? Where would the data, applications, and users come from?

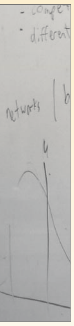
One answer came from the then-emerging world of Web 2.0. Although very different from the Semantic Web in its origins, focus, and aspirations, Web 2.0 has created several interesting opportunities to light up the Semantic Web. For example, it brought an entirely new kind of data that could be captured from many of the social-networking sites that arrived on the Web in 2003. Studying social networks was hardly new. The field of social-network analysis existed for most of the 20th century, producing a distinct set of analysis tools and many interesting results. However, these tools were designed to study relatively small data sets obtained by carrying out complete surveys in small, real-world communities such as school classes, offices, and villages.

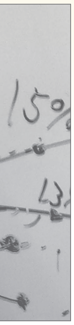
Clearly, the setting of Web-based social networks was very different. The data sets we found were much larger than those studied by previous survey methods. Also, the data was more fragmented, and it wasn't even clear whether the relationships noted were meaningful representations of real-world relationships. Together with my advisors Hans Akkermans, Tom Elfring, and Peter Groenewegen, we've studied methods to extract, represent, and reason with social-network data sets obtained from the Web. In our particular case study, we've collected data about researchers in the Semantic Web community and have successfully used the data set to predict researchers' performance on the basis of their social networks' properties.

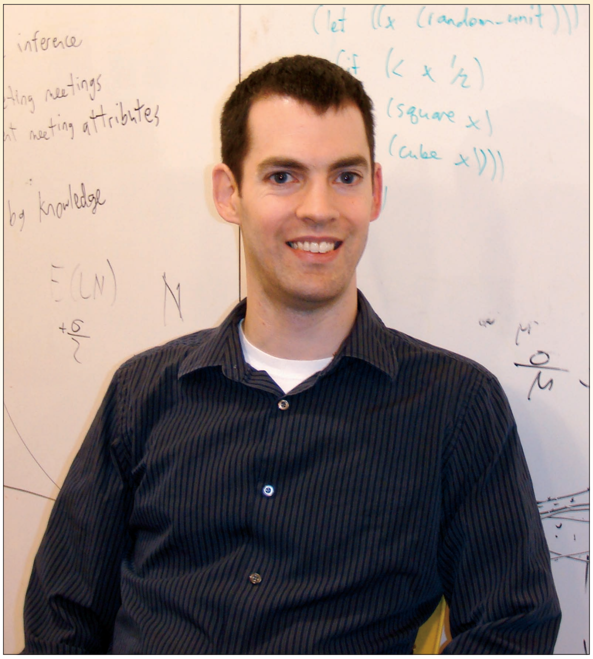
I also noted that some Web 2.0 systems had made progress in realizing one of the Semantic Web's original aspirations: motivating everyday users to annotate the Web. Tagging had become a popular way to organize content on Web 2.0 sites, where most of the content was provided by the users themselves and therefore could best be described by those same users. Tagging systems have become the source of a very interesting kind of semantics: the semantics obtained by observing tag usage. My contribution to the study of these *folksonomies* was their formal representation and the use of network-analysis methods to extract their semantics.

My current research at Yahoo! focuses on realizing one of the most promising applications of Semantic Web technology: improving information retrieval on the Web. As my colleagues and I recently explained in a Trends & Controversies article in this magazine (Jan./Feb. 2008), semantic search is a grand challenge requiring the combined effort of the information-retrieval, NLP, and Semantic Web communities—communities that traditionally have been separated, much like the social science and computer science communities. However, the potential benefits are immense: semantic search holds the promise of a qualitatively different search experience based on query-driven information integration across the Web, which we believe will impact how millions of users interact with the Web every day.









Brian Milch

MIT Computer Science and Artificial Intelligence Laboratory

Brian Milch is a postdoctoral researcher at the Massachusetts Institute of Technology's Computer Science and Artificial Intelligence Laboratory. His research focuses on inference and learning for models that combine logical and probabilistic forms of knowledge representation. Milch received his BS with honors in symbolic systems from Stanford University and spent a year as a Google research engineer before earning his PhD in computer science from the University of California, Berkeley. He has received a US National Science Foundation Graduate Research Fellowship and a Siebel Scholarship, and his PhD dissertation was nominated for the ACM Doctoral Dissertation Award. When not at work, he enjoys swimming, reading, and keeping up with current events; he also recently learned to scuba dive. Contact him at milch@csail.mit.edu.

Combining Logic and Probability

My research revolves around combining probabilistic and logical approaches to knowledge representation, reasoning, and learning. Intelligent systems often need to draw inferences from noisy or ambiguous data; probability serves as a principled framework for weighing possible hypotheses.

The development of graphical models such as Bayesian networks, which compactly specify the dependencies among large sets of random variables, has led to successful probabilistic systems for a wide range of tasks. However, probabilistic network models lack the expressive power of first-order or relational logic. For instance, a Bayesian network can say that John's choice of clothing depends on the weather in Boston and that Marie's choice of clothing depends on the weather in Paris. However, it can't express the general rule underlying these dependencies: each person's choice of clothing depends on the weather where he or she lives.

The need to learn and reason with probabilistic models that generalize across objects—analogous to quantified formulas in a logical knowledge base—motivates the growing field of *statistical relational learning*. My thesis work introduced a relational probabilistic modeling language called Bayesian logic, or BLOG, which goes beyond earlier languages in representing scenarios in which the relevant objects are initially unknown. Such scenarios are ubiquitous in practice. For instance, if an information extraction system finds some documents with information about "Thomas Smith" and others with information about "Tom Smith," it must determine how many distinct people are being referred to.

Although many special-purpose algorithms have addressed this problem of "data association," "record linkage," or "coreference resolution," BLOG makes it possible to reason about unknown objects within a general-purpose modeling formalism. Performing inference on probabilistic models with large numbers of objects is a major challenge. Our inference engine for BLOG uses a Markov chain Monte Carlo algorithm, performing a random walk over "possible worlds" with varying sets of objects.

With my MIT colleagues, I'm investigating inference algorithms that gain efficiency by reasoning about whole groups of interchangeable objects at once. This research is inspired by techniques from logical theorem proving; it exemplifies an exciting trend toward the integration of probabilistic and logical inference. I'm also working on algorithms for learning relational probabilistic models from data. In the long term, I hope to build systems that not only learn dependencies between variables but also invent new predicates. For instance, a system looking at the author lists of papers could explain coauthorship patterns by hypothesizing a "collaboration" predicate that holds between certain researchers.

The logical and probabilistic branches of AI have taken largely separate paths over the past 20 years. Combining the strengths of these two approaches promises to enable a broad range of practical systems and to move us forward on the path toward human-level AI.

Multimodal Perception of Human Nonverbal Behaviors

I envision a world where

- a car can recognize that a driver isn't paying attention and alert him or her about a possible incoming collision,
- a virtual human can welcome you into a new office or exhibition and direct you to the right location, and
- computers can analyze a patient's nonverbal behaviors and help doctors assess his or her psychological condition.

For this world to happen, we must provide computer interfaces with the ability to recognize, understand, and respond to people's nonverbal behaviors. This is the focus of my work.

I call my research area *multimodal perception* because it combines various communication channels—spoken words, intonation, and visual gestures—allowing computers to recognize and understand the context of people's gestures and emotions. The pace at which a person speaks, where they fix their gaze, when they are quiet—all these factors affect the timing and meaning of a listener's visual feedback, such as nodding a head and making or breaking eye contact. Humans integrate all this information automatically, but computers must be taught. My research facilitates this by identifying people's natural nonverbal conversational cues and, most importantly, developing efficient, robust algorithms for computers to comprehend and utilize these visual gestures.

An important milestone of my research career was to create a new mathematical model (Latent-Dynamic Conditional Random Fields) that could explicitly model the substructure of a specific communication channel as well as the interaction between different modalities. This technique has outperformed previous state-of-the-art approaches for recognition of nonverbal behaviors, such as hidden Markov models, support vector machines, and conditional random fields. I also developed Watson, a real-time library for visual-feedback recognition used at hundreds of labs and institutions worldwide. Watson helps robots and virtual humans not only identify movement such as head gestures and eye gaze but also anticipate when a gesture will likely happen, allowing these computer interfaces to appear more human in their communications.

Multimodal perception is a multidisciplinary research topic overlaying the fields of computer vision, human-computer interaction, social psychology, machine learning, and AI. Applications include areas as diverse as robotics, education, and entertainment. Future research directions in this domain involve automatically selecting only the most relevant information from the current dialogue context, finding the optimal representation for these contextual features, and extending current mathematical models to efficiently fuse the different information sources.

In my vision of AI's future, computer interfaces will be able to efficiently react to a confused user, keep up with natural turn-taking dynamics, and smoothly perceive a user's state of mind (for example, agreement or disagreement). Such interfaces will enable a more efficient, engaging interaction between robots and people.

AI's 10 to Watch

Louis-Philippe Morency

USC Institute for Creative Technologies

Louis-Philippe Morency is a research scientist at the University of Southern California's Institute for Creative Technologies, where he leads the Nonverbal Behaviors Understanding project. His main research interest is multimodal perception of human nonverbal behaviors, integrating various channels of communication—semantic, prosody, and visual cues—to recognize and understand the context of people's gestures and emotions. He developed Watson, a real-time library for nonverbal-behavior recognition. Watson became the de facto standard for adding perception to embodied agent interfaces and has been downloaded by more than 100 researchers worldwide. Morency received his BS in computer engineering from Laval University and his MS and PhD in computer science from the Massachusetts Institute of Technology. He received the International Conference on Multimodal Interfaces Best Paper Award for two consecutive years for his research on context-based recognition and multimodal adaptation. Contact him at morency@ict.usc.edu.

AI and Ontology Technologies

Ontologies are vocabularies of terms often developed collaboratively by a community of users. With the increasing importance of ontologies in academia and industry, the development of ontology languages and efficient tools supporting ontology-based applications has become a main research goal of the Semantic Web and knowledge representation communities.

It is now widely agreed that ontology languages should be based on formal logic. Thus, knowledge representation formalisms called *description logics* (DLs) are used to provide the formal underpinning for the Web Ontology Language (OWL)—the family of ontology languages used in the Semantic Web. Using DLs has several important advantages. On the theoretical side, DLs provide OWL with a semantic framework that has well-understood formal properties. On the practical side, OWL users can reuse tools and reasoners developed by the DL community.

OWL has been very successful in the past few years; however, the practical experience has raised numerous exciting and challenging research questions. Many of these involve extending the expressive power of OWL while retaining the good computational properties of the language such as decidability of reasoning. Much research has also focused on improving the scalability of reasoning techniques. Finally, researchers have noted that, in some applications, there is a gap between the users' intuition and the modeling style enforced by OWL. This is particularly true in data-centric applications, which suggests that rethinking the fundamentals of OWL in the light of relational and object-oriented databases might be necessary.

My research has tried to address some of these problems. In my PhD work, I developed several reasoning algorithms with the goal of applying optimization techniques from deductive databases to OWL reasoning. I have implemented them in the KAON2 reasoner, which has proven effective in data-centric applications and has found its way into commercial use through ontoprise GmbH, a Karlsruhe-based company. In my postdoctoral research, I worked on integrating OWL with nonmonotonic formalisms such as logic programming, bringing OWL closer to schema languages used in relational databases, and extending OWL with capabilities for representing objects with complex structures. Recently, my collaborators and I have developed another reasoning algorithm for OWL based on hypertableau—a theorem-proving calculus that is, in many cases, more efficient than the commonly used tableau-based OWL reasoning algorithms. I have implemented this algorithm in the HermiT reasoner, which allowed me to classify complex ontologies such as certain versions of the GALEN biomedical terminology—a long-standing open problem for OWL reasoning.

Ontology-related research will likely remain active for some time to come. The scalability of ontology reasoning will continue to pose theoretical and practical challenges. Bringing ontology-based technologies into the mainstream of computer science and industry will require a multidisciplinary effort, integrating experts with knowledge about diverse application domains, theoretical computer science, and practical implementation techniques.

AI's 10 to Watch

Boris Motik

Oxford University Computing Laboratory

Boris Motik is a research associate at the Oxford University Computing Laboratory. He completed his bachelor's and master's studies at the University of Zagreb in Croatia, after which he worked for several years as a software developer in Croatia, the US, and Germany. He then returned to academia and completed his doctoral studies at the University of Karlsruhe under the supervision of Rudi Studer. He continued his postdoctoral studies in Manchester and subsequently at Oxford. His paper "On the Properties of Metamodeling in OWL" won the Best Paper Award at the 2005 International Semantic Web Conference. He also received the 2007 Cor Baayen Prize, given by the European Research Consortium for Informatics and Mathematics to "a most promising young researcher in computer science and applied mathematics." Contact him at boris.motik@comlab.ox.ac.uk.

Jennifer Neville

Purdue University

Jennifer Neville is an assistant professor at Purdue University with a joint appointment in the Departments of Computer Science and Statistics. Her research interests lie in statistical relational learning, which is concerned with the development of data mining and machine learning techniques for complex relational domains. Her research focuses on the development and analysis of models that can exploit statistical dependencies among related instances, as well as algorithms to accurately learn those models. Applications of her work include social-network analysis, fraud detection, citation analysis, and bioinformatics. Neville received her PhD in computer science from the University of Massachusetts Amherst, where she was a member of the Knowledge Discovery Laboratory. She has been selected as a DARPA IPTO (Information Processing Techniques Office) Young Investigator and has served as a member of the DARPA Computer Science Study Panel. Contact her at neville@cs.purdue.edu.

AI's 10 to Watch

Combining Logical and Statistical AI

Recently, researchers in statistical relational learning have combined techniques from AI, databases, and statistics to model complex relational data sets such as social networks, the World Wide Web, and protein-protein interaction graphs. This research has significantly extended the range of concepts that can be accurately learned with statistical models. It has also led to what some have termed a "relational revolution" in the field of automated learning and discovery by moving beyond the long-held assumption of independent and identically distributed examples.

The power of relational data lies in combining intrinsic information about objects in isolation with information about related objects and the connections among those objects. Consider the task of identifying stockbrokers involved in securities fraud. The US National Association of Securities Dealers (NASD) currently identifies potentially fraudulent brokers using a set of handcrafted rules that evaluate broker attributes in isolation. However, relational information is often considered central to fraud detection because fraud and malfeasance are social phenomena, communicated and encouraged by the presence of other individuals who also wish to commit fraud. Relational-learning techniques can automatically identify patterns in the professional and organizational relationships among brokers, which are likely to be more accurate indicators of fraud than the handcrafted rules. The NASD could therefore use these techniques to better target its limited regulatory resources on brokers that are more likely to engage in fraudulent behavior.

My research has focused on the design and analysis of statistical models for such relational domains, along with algorithms for learning the structure and parameters of those models. In particular, I've developed efficient, yet accurate, approximate learning algorithms to make relational analysis feasible for large-scale relational data sets where other relational approaches are computationally intensive, if not intractable. Notably, these models are the only methods to date that adjust for statistical biases toward particular features that occur owing to relational data's unique characteristics. In addition, I've developed a framework for experimental analysis of relational models that decomposes error due to both the relational-learning process and the collective-inference process, which is a new source of error. Initial analysis in this framework has improved our understanding of a broad class of relational algorithms.

My current research focuses on developing methods for dynamic, transactional, and multisource networks—where the relationships change over time, the observed relationships are of low granularity, and/or the relational structure is partially observed from different sources. Although researchers have recently produced several successful relational models, this research has focused primarily on attribute prediction in static, well-structured, fully observed domains. To date, there are few techniques for analyzing real-world relational domains that don't meet these assumptions (for example, online social networks and intelligence analysis).

Statistical relational learning is still in its early stages. However, its widespread success indicates that an interdisciplinary approach to model and algorithm development might be the key to making fundamental advances in AI.

AI's 10 to Watch

Erik B. Sudderth

University of California, Berkeley

Erik B. Sudderth is a postdoctoral scholar in the Department of Electrical Engineering and Computer Science (EECS) at the University of California, Berkeley. In 2009, he will join the faculty of the Brown University Department of Computer Science. His research interests include probabilistic graphical models; nonparametric Bayesian methods; and applications of statistical machine learning in image processing, tracking, object recognition, and visual scene analysis. Sudderth received his bachelor's degree (summa cum laude) in electrical engineering from the University of California, San Diego, and his master's and PhD in EECS from the Massachusetts Institute of Technology. His dissertation, which was jointly supervised by William Freeman and Alan Willsky, received honorable mention for MIT's George M. Sprowls Award. Erik has received a US National Defense Science and Engineering Graduate Fellowship and an Intel Foundation Doctoral Fellowship. Contact him at sudderth@eecs.berkeley.edu.

Learning Representations for Visual Scenes

Visual data provides an incredibly rich source of information about the world around us. From a casual glance at a scene, we can recognize and categorize objects, estimate their 3D shape, and even determine the materials composing them. My computer vision research explores new approaches to automated scene interpretation, which draw on statistics and machine learning advances. I believe that careful analysis of the statistical relationships underlying realistic images leads to more robust, effective vision algorithms.

Although we have no trouble recognizing a new model of automobile or a stranger's face, similarly reliable computer-based object categorization isn't currently possible. This task is particularly challenging owing to the wide variability within many categories (consider, for example, the differences among dog breeds or chair styles). Traditional machine learning methods can be effective given very large, hand-labeled training sets, but they don't easily scale to the thousands of categories found in natural scenes. My doctoral research adapted probabilistic graphical models to learn hierarchical representations of object appearance. By designing an integrated model that transfers knowledge among multiple object categories, we can build systems that better generalize from a small set of labeled training examples.

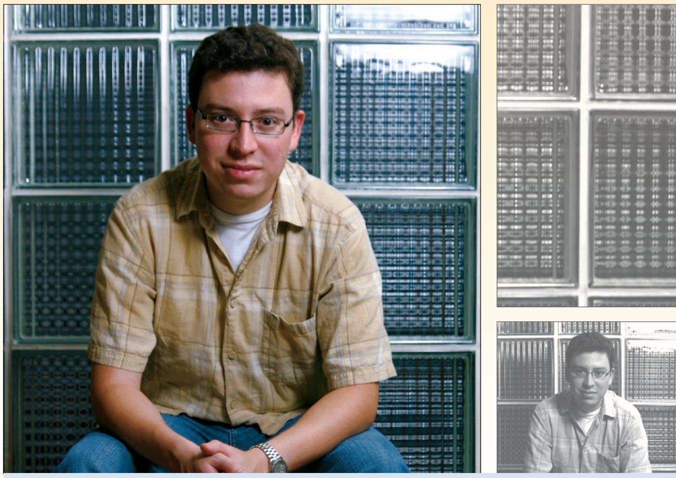
Although probabilistic models have a rich history in computer vision, their performance has been limited by the use of overly simplified, hand-designed caricatures of real scenes. To improve robustness, my colleagues and I have explored methods for discovering visual relationships directly from training images. Nonparametric Bayesian statistical methods reduce sensitivity to prior assumptions by adapting infinite-dimensional models. When few observations are available, they favor simple predictions based on a small, easily estimated set of latent variables. However, their flexible form leads to data-driven learning algorithms that capture additional detail in large, complex data sets. My research has explored simulation-based Markov chain Monte Carlo methods, variational methods with rich connections to optimization theory, and hybrid learning algorithms that combine these methods' best features. One continuing goal is to understand when, and why, such approaches are effective.

Applying these nonparametric Bayesian methods, I've developed hierarchical generative models for objects, the parts composing them, and the scenes surrounding them. These models allow an unknown set of objects, and their corresponding internal structure, to be efficiently inferred from training images. Going forward, such methods seem ideally suited to leverage the rapidly increasing availability of large, partially annotated image and video databases. One ongoing project uses a data set of manually segmented images to validate the statistical biases underlying nonparametric prior distributions.

By incorporating knowledge about contextual relationships and global scene structure, hierarchical models blur the boundaries between traditional vision tasks such as segmentation, tracking, and recognition. I anticipate future flexible vision systems that adapt as they observe new images and that transfer knowledge among tasks as diverse as multimedia retrieval, document analysis, and robotic navigation. By exploring the complexities of visual scenes, we might also attain more general insights into human and artificial intelligence.

Luis von Ahn

Carnegie Mellon University



AI's 10 to Watch

Luis von Ahn is an assistant professor in Carnegie Mellon University's Computer Science Department. In 2006, his first year on the faculty, he received a MacArthur Fellowship and was chosen one of *Popular Science* magazine's Brilliant 10. More recently, he was named to Silicon.com's 50 most influential people in technology, *Smithsonian* magazine's America's Top Young Innovators in the Arts and Sciences, and *Technology Review*'s TR35: Young Innovators under 35. Von Ahn received his BSc in mathematics from Duke University and his MSc and PhD in computer science from CMU under advisor Manuel Blum. His thesis on human computation earned the Best Doctoral Dissertation Award from CMU's School of Computer Science. He also received the school's Alan J. Perlis Student Teaching Award and the Herbert A. Simon Award for Teaching Excellence in Computer Science. His research interests include enticing people to do work for free, as well as catching and thwarting cheaters in online environments. Contact him at biglou@cs.cmu.edu.

Human Computation

I'm working on *human computation*, a new area of computer science that harnesses the combined computational power of humans and computers to solve problems together that neither can solve alone. Tasks such as image recognition are easy for humans but still confound even the most sophisticated computational techniques. Traditional approaches focus on improving algorithms. In contrast, I'm pursuing an approach that uses compelling methods, such as games, to entice people to participate in collective computations that solve large-scale open problems.

This approach is exemplified by my collaborative ESP Game (www.espgame.org), in which people label images on the Web as a side effect of playing. The ESP Game presents Web images to randomly matched pairs of players, who type possible one-word descriptions. When two partners type the same word, that word becomes a label for the image. Through various techniques, the game ensures that the labels are meaningful and extremely accurate, even if the players don't want them to be. The ESP Game has collected more than 50 million labels. These labels can improve Web image search, increase Web site accessibility for visually impaired individuals, and help Web browsers block pornography.

My students and I are developing additional games to address other important open problems in AI: language translation, text summarization, Web search improvement, and annotation of sound and video clips. These "games with a purpose" (GWAPs) have many applications, such as generating training data that can help improve automated methods for solving these problems.

Although such games represent the first seamless integrations of gameplay and computation, we're left with the question of how generally useful this approach is. My ongoing research in this area will focus on defining the concept of GWAPs, developing a general process for constructing them, and enabling and encouraging the rest of the research community to participate in this paradigm.

Another example that leverages human processing power is my research on CAPTCHAs—automated tests that humans can pass but computer programs can't. You've probably encountered them on Web registration forms: images of distorted text that you must type correctly to register for an account, purchase tickets, and so on. Many Web sites use CAPTCHAs to ensure that only humans obtain free email accounts and other services. People type an estimated 200 million CAPTCHAs every day, each taking roughly 10 seconds. My reCAPTCHA project (www.recaptcha.net) attempts to channel these millions of hours of human effort into "reading" books online.

Multiple projects are underway to digitize physical books: pages are photographically scanned, then converted into ASCII text using optical character recognition (OCR) to make the books searchable. Although OCR achieves extremely high accuracy on newer texts, many books scan poorly due to age, quality, or damage (for example, pencil markings). In such cases, OCR achieves a low recognition rate.

ReCAPTCHA improves digitization by sending the words that OCR can't read to the Web in the form of CAPTCHAs for humans to decipher. To determine whether humans are entering the correct answers, reCAPTCHA uses a control word. So, the program gives each new word that OCR can't read correctly to a user, along with a second image of a word for which we already know the answer. ReCAPTCHA then asks the user to type both words. If the user correctly types the one for which the answer is already known, we assume their answer for the new one is also correct. The program then gives the new image to other people to determine with higher confidence whether the original answer was correct. Using this technique, we're digitizing more than five million words per day.