AAAI-07 Workshop Reports

Sarabjot Singh Anand, Daniel Bahls, Catherina R. Burghart, Mark Burstein, Huajun Chen, John Collins, Tom Dietterich, Jon Doyle, Chris Drummond, William Elazmeh, Christopher Geib, Judy Goldsmith, Hans W. Guesgen, Jim Hendler, Dietmar Jannach, Nathalie Japkowicz, Ulrich Junker, Gal A. Kaminka, Alfred Kobsa, Jerome Lang, David B. Leake, Lundy Lewis, Gerard Ligozat, Sofus Macskassy, Drew McDermott, Ted Metzler, Bamshad Mobasher, Ullas Nambiar, Zaiqing Nie, Klas Orsvarn, Barry O'Sullivan, David Pynadath, Jochen Renz, Rita V. Rodriguez, Thomas Roth-Berghofer, Stefan Schulz, Rudi Studer, Yimin Wang, and Michael Wellman

■ The AAAI-07 workshop program was held Sunday and Monday, July 22–23, in Vancouver, British Columbia, Canada. The program included the following thirteen workshops: (1) Acquiring Planning Knowledge via Demonstration; (2) Configuration; (3) Evaluating Architectures for Intelligence; (4) Evaluation Methods for Machine Learning; (5) Explanation-Aware Computing; (6) Human Implications of Human-Robot Interaction; (7) Intelligent Techniques for Web Personalization; (8) Plan, Activity, and Intent Recognition; (9) Preference Handling for Artificial Intelligence; (10) Semantic e-Science; (11) Spatial and Temporal Reasoning; (12) Trading Agent Design and Analysis; and (13) Information Integration on the Web.

Acquiring Planning Knowledge via Demonstration

In the mid to late 1980s there was a flurry of papers using explanation-based techniques to learn how to perform complex actions by observing (or interpreting descriptions of) human performance. These techniques were shown to work reasonably well with one or a small number of examples. However, as statistical approaches gained in power and popularity, and some kinds of data became more plentiful, machine learning as a field moved away from this kind of learning, which requires strong domain models. Recently, efforts have begun to look again at explanation-based learning and other approaches in contexts where few examples can be gathered. This workshop focused specifically on techniques for learning planning and procedural knowledge from single demonstrations, a task made more difficult by the fact that some contingency situations and what to do about them are never demonstrated.

The guest speaker was James Allen, who gave an extended talk on the procedural learning on the web (PLOW) system that was also described in his team's outstanding award-winning paper during the main conference. PLOW learns by combining observation of user-demonstrated web procedures with an interpretation of the user's natural language narration of that procedure. This enables learning from one example by filling in critical details about such things as the criteria for decisions about alternate courses of action, when to end loops, and the semantic relationships among parameters to service calls. A similar set of themes was echoed in the talk of James Blythe describing ISI's Tailor system. Here the natural language instruction was replaced by a more structured user interface in order to learn information-gathering procedures. Walsh and Littman also described the modeling, planning, and execution algorithm they call MOPLEX, a system that learns a conceptual model of information query procedures.

Mark Burstein presented an overview of plan order induction by reasoning from one trial (POIROT), a system being developed as part of DARPA's Integrated Learning program that uses a shared blackboard of hypotheses generated by different learning systems observing the same demonstration as the basis for its procedure learning. By combining the hypotheses produced by more bottom-up, inductive techniques such as that presented by Tim Oates and Fusun Yaman of the University of Maryland, Baltimore County on the web initiative in teaching (WIT) system, and more top-down explanation-based techniques found in LIGHT, a system from Stanford's ISLE, POIROT is able to

produce reasonably complete procedures from one example. The effectiveness of these procedures is tested by CMAX, a component that designs experiments to test the completeness and consistency of generated hypotheses. Clayton Morrison of ISI described this system for the assembled group.

A number of other talks, motivated in part by their roles in the generalized integrated learning architecture (GILA) system, another project in the Integrated Learning program, described aspects of procedure learning associated with learning to select appropriate values for real-valued parameters in actions that form parts of complex procedures in physical domains. One talk, presented by Ugur Kuter of the University of Maryland representing the larger group, described a system called ConstraintLearner that learns safety constraints on parameter assignments in developed plans for the target action domain from a user demonstration plus domain knowledge of absolute parameter bounds. A group from Oregon State presented a paper on the optimization of learned cost functions for action selection from similar extended demonstrations. The model learns cost functions that capture the expert demonstrator's preferences over features of the goal state. Sungwook Yoon and Rao Kambhampati presented a paper on an approach built around a hierarchy of hybrid representations, with actions at the higher level represented as symbolic encapsulations (turn right, move to position *p*) of policies, learned from user-tagged demonstrations, and actions at the lower level represented as optimizing value functions, primarily for computing parameters by mimizing cost.

The day ended with a lively discussion of the role of background knowledge in these systems and some disagreements over the role of multiple and potentially inconsistent representations in these hybrid learning systems.

The papers from the workshop were published as AAAI Technical Report WS-07-02 and are available from AAAI Press.

—Mark Burstein, Jim Hendler, Tom Dietterich, and Drew McDermott

Configuration

Representing and solving configuration problems have always been subjects of interest for the artificial intelligence community. Powerful knowledge representation formalisms are necessary to capture the great variety and complexity of configurable products. Furthermore, efficient reasoning methods are required to provide intelligent interactive behavior in configuration software, such as solution search, satisfaction of user preferences, personalization, optimization, explanation, diagnosis, and so on.

Nowadays, different AI approaches are well established as central technologies in many industrial configuration systems. This widespread industrial use of AI-based configurators makes the field more challenging than ever: the complexity of configurable products is still growing, the mass-customization paradigm is extending to fields like web service and software configuration, personalized (web-based) user interaction and user preference elicitation are of increasing importance, and finally, the integration of configurators into surrounding information techology infrastructures, such as business information systems or web applications, has become a critical issue.

This workshop continued the series of successful configuration workshops started at the AAAI 1996 Fall Symposium and continued at IJCAI, AAAI, and ECAI since 1999. As well as researchers from a variety of different fields, the configuration workshop series has also attracted a significant number of industrial participants from major configurator vendors like Tacton, SAP, Oracle, and ILOG, along with end users like Siemens, HP, and Daimler-Chrysler.

The primary goal of the workshop was to promote high-quality research in all technical areas related to configuration. As such, the workshop was of interest to researchers working in the various fields within the wide range of applicable AI technologies (for example, constraint reasoning, description logics, nonmonotonic reasoning, case-based reasoning, and so on). It served as a platform for researchers and industrial participants to exchange needs, ideas, benchmarks, and case studies. Colocated with AAAI 2007, the Configuration workshop provided an ideal forum to attract high-quality submissions.

The workshop ran over one and a half days and comprised eight technical presentations addressing such topics as collaborative and web services for configuration, agent-based configuration, and constraint-based reasoning and preferences for configuration.

Two invited talks were also given. Klas Orsvarn from Tacton Systems AB (Sweden) gave a talk titled "Tacton Configurator—A Comprehensive Constraint-Based Configurator." A very special feature of this talk was a detailed demonstration of the Tacton configurator. Ulrich Junker from ILOG SA (France) also gave a talk, titled "Configuration as Unbounded Hierarchical Decision Making: How can Users Find an Easy Way through Complex Configurations?" His talk addressed a number of important aspects of preference elicitation and reasoning that are relevant in large real-world configuration contexts in which products are very complex.

A tribute was paid to the late Pragnesh Jay Modi (Drexel University, USA), who passed away in April of 2007 and who was a coauthor of a paper presented at the workshop. Pragnesh's death was not only a huge personal loss to his family and friends but also a major loss to the field of artificial intelligence.

The papers from the workshop were published as AAAI Technical Report WS-07-03 and are available from AAAI Press.

—Barry O'Sullivan and Klas Orsvarn

Evaluating Architectures for Intelligence

Architectures form an integral part of artificially intelligent agents and robots. Architectures structure and organize the knowledge used by the agents to select actions in dynamic environments, plan and solve problems, learn, and coordinate with others. Architectures serve to integrate general capabilities expected of an intelligent agent (for example, planning and learning), to implement and

test theories about agent cognition, and to explore domain-independent mechanisms for intelligence.

As AI research has improved in formal and empirical rigor, traditional evaluation methodologies for architectures have sometimes proved insufficient. Formal analysis has often proved elusive; we seem to be missing the notation required for proving properties of architectures. Experiments that demonstrate generality are notoriously expensive to perform and are not sufficiently informative. And at a high level, evaluation is difficult because the criteria are not well defined: Is it generality? What is the ease of programmability? Is it compatible with data from biology and psychology? There are no established evaluation methodologies and only a handful of established evaluation criteria.

Recognizing that scientific progress depends on the ability to conduct informative evaluation (by experiment or formal analysis), this workshop addressed the methodologies needed for evaluating architectures. The focus was on methodology, rather than specific architectures. This workshop's goal was to propose and discuss evaluation criteria for architectures, with the hope of generating a concrete result that will initiate a real process within the community.

Over two days, presentations raised challenges, addressed questions, and tackled the evaluation challenge from a variety of perspectives. Contributed and invited talks addressed formal approaches, benchmarking, challenge problems, competitions, and test beds. Several talks reflected on related past efforts and attempted to draw lessons.

Based on presentations and moderated discussions, two different results of the workshop were achieved. First, it became clear that the community actually mixes three different goals in investigating architectures: (1) modeling human behavior and cognition; (2) generating intelligent behavior, whether psychologically valid or not; and (3) more easily solving specific tasks, or building systems, in which the use of architecture eases development. Second, for each of these purposes, different methodologies can be used for evaluation, although some lend themselves more closely to specific goals. For instance, compatibility with psychological data sets is very important if the goal of the architecture is to model human behavior and cognitive processes but is of almost no importance when trying to address a specific task requiring intelligence. Competitions and benchmarks are more useful for addressing specific tasks, or for general intelligence. Formal methods seem to be needed in almost all areas. Most proposals and presented methodologies could be categorized along these axes of classification.

The papers from the workshop were published as AAAI Technical Report WS-07-04 and are available from AAAI Press.

—Gal A. Kaminka and Catherina R. Burghart

Evaluation Methods for Machine Learning

This year's workshop continued the discussion engaged at last year's AAAI workshop on the same topic, which had concluded that there are serious drawbacks in the way in which we evaluate learning algorithms. In particular, the participants of last year's workshop had agreed on two points. First, that our evaluation practices are too narrow: other properties of algorithms (for example, interpretability, performance under changing conditions, evaluation of the transfer mechanism, and so on) should be tested in addition to their accuracy, which, itself, should be considered more flexibly; and second, that the University of California, Irvine (UCI) datasets do not reflect the variety of problems to which learning algorithms are applied in practice. The invited talks at this year's workshop were designed to address these two issues.

Regarding the first issue, Yu-Han Chang gave an interesting talk that considered evaluation metrics designed to measure the performance of transfer learning algorithms. Rich Caruana presented his investigation of various evaluation metrics, showing that what matters more than the choice of a metric is the way in which that metric is being used. He showed that, in fact, without proper calibration the results we obtain from certain classes of metrics could be meaningless.

To address the second issue, we invited two researchers that do not develop new machine-learning algorithms but rather specialize in their application to specific areas. In particular, George Tzanetakis presented his research in audio processing and music information retrieval while Andre Kushniruk shared the lessons he learned about evaluation from his study of emerging health-care information systems. Both agreed that the issues encountered in their respective real-life domains go well beyond those considered by the UCI domains and that our evaluation practices fall short of what is really necessary in their respective fields. For example, Kushniruk found that no performance measures were available for the kind of testing he was interested in (for example, a measure of low-cost portable usability), while Tzanetakis deplored the fact that we do not "listen to our data" enough and explained how commonly used procedures such as cross-validation could not be applied to his field without serious prior thought as to what exactly they compute and what the consequential results really mean. In additions to these talks, nine papers were presented. A number of these, like Rich Caruana's invited talk, took a high-level view of performance metrics, pitting them against one another. Alexander Liu, for example, presented a framework for analyzing skew in evaluation metrics. Alexandru Niculescu-Mizil discussed the kind of issues arising when different performance measures are used at design and deployment time. Payam Refaeilzadeh took a similar direction, but with respect to a different aspect of classifier evaluation: rather than metrics, he compared two different cross-validation regimens for the comparative evaluation of feature selection algorithms. In the same vein, and following up on last year's suggestion to concentrate on visualizing evaluation, Nathalie Japkowicz presented a projection method that allows for a quick visualization of several evaluation metrics simultaneously. Finally, Eduardo Costa reviewed a number of performance evaluation measures for hierarchical classifiers.

Two papers presented studies on practical domains and discussed their experience with existing evaluation metrics. In particular, Dympna O'Sullivan presented her work on the prediction of pediatric asthma, and Robert Schrag explored the similarity between the scoring of hypotheses from threat-detection technologies and machine-learning evaluation.

Finally, the paper by Chris Drummond titled "Making Evaluation Robust, but Robust to What?" started a very exciting and lively debate as to what robustness is, in the context of classification, and how it could be measured. The workshop ended with a panel discussion that was very successful at drawing in comments from the audience. Additional questions raised in this part of the workshop included: Is there a need for machine-learning researchers to be more fluent in statistical theory? Should we be aiming at a generalized framework for evaluation, or should evaluation methods be designed on a case by case or application domain by application domain basis? How should our basic evaluation methods (cross-validation, statistical testing) be used to ensure proper testing? Is the field of machine learning progressing, or have we reached a plateau? Can our current evaluation methods answer this question? What are our workshops achieving? Is there a need for an additional workshop next year?

The cochairs of this workshop were Chris Drummond (NRC Institute for Information Technology), William Elamzeh (University of Ottawa), Nathalie Japkowicz (University of Ottawa), and Sofus Macskassy (Fetch Technologies). The papers from the workshop were published as AAAI Technical Report WS-07-05 and are available from AAAI Press.

—Nathalie Japkowicz, Chris Drummond, William Elazmeh, and Sofus Macskassy

Explanation-Aware Computing

Transparency and trust play an important role in the acceptance of information systems. Systems able to explain their decisions, concepts, and information sources to users can demonstrate their trustworthiness and support users' understanding. In addition, systems able to generate explanations for their own internal use and to reason about their explanations may be able to increase their robustness in dealing with unexpected situations, as well as to improve their future performance, by using explanations to refine their internal models and reasoning processes. All of these benefits depend on the ability of the systems to generate high-quality explanations and to exploit explanations in their processing.

Advancing the capability of systems to generate and use explanations depends on advances in the models, methods, and tools available for managing explanation-relevant information, as well as on developing effective methods for retrieving explanation-relevant information and integrating explanation and application knowledge. Beyond technical considerations, the design of effective explanation systems must reflect fundamental insights about the nature and use of explanations, as illuminated, for example, in philosophical and psychological investigations, as well as by social perspectives on the context and use of information technology applications.

Disciplines such as artificial intelligence, cognitive science, linguistics, philosophy of science, and education have all considered varying aspects of explanation. Each has insights to share, but few opportunities exist for interaction and clarification of their different views. This AAAI workshop, known as ExaCt 2007, provided such a forum. This two-day workshop, a sequel to the AAAI Fall Symposium (ExaCt 2005), brought together researchers and practitioners from diverse groups to study explanation issues and explore the requirements for effective explanation in information technology applications.

The 10 papers presented at the workshop reflected a wide variety of approaches, and the workshop program was designed to foster consideration of central explanation issues from different perspectives. The first day of the workshop opened with a philosophical view, an invited talk by Douglas Walton, professor of philosophy at the University of Winnipeg, on dialogical models of explanation. The second day began with an AI view, an invited talk by Bruce Porter, professor of computer science at the University of Texas, Austin, on developing a new class of task-independent knowledge systems that are designed to be able to answer a broad range of questions in a domain and the challenges such systems pose for explanation. Substantial time was reserved for discussions, which were far ranging, with topics including the relationships between explanation, justification, and provenance and the trade-offs between fully automated explanation construction and user-guided navigation through explanatory structures. Discussion and debate were vigorous and productive throughout, and at the close of the workshop, the attendees enthusiastically supported the organization of a future meeting to continue the dialogue. Planning for this meeting has begun.

The cochairs of this workshop were Thomas Roth-Berghofer (TU Kaiserslautern), Stefan Schulz (The e-Spirit Company), and David Leake (Indiana University). The demonstrations chair was Daniel Bahls (DFKI). The papers from the workshop were published as AAAI Technical Report WS-07-06 and are available from AAAI Press.

—Thomas Roth-Berghofer, Stefan Schulz, David B. Leake, and Daniel Bahls

Human Implications of Human-Robot Interaction

The international and multidisciplinary gathering at this workshop sensed timeliness and importance in its topic. Human-robot interaction (HRI) already has established itself as a domain with substantive "how to" problems for technical communities, and it appears that these problems can involve open questions regarding humans as well. Moreover, technological advances in areas such as humanoid service robotics increasingly are drawing attention within the humanities and social sciences. Artifi-

cially intelligent artifacts displaying convincing humanlike behavior and appearance tend to be viewed readily as more than "tools"—they become "mirrors" reflecting distinctively human issues of ethics, personhood, privacy, and the like.

The workshop's opening presentation, for example, investigated "consequences for human beings" of machine ethics research—particularly, research supporting creation of ethical robots. Noting a need for this kind of research in applications such as robotic elder care, the authors identified other human benefits that might be expected from the work. One of these benefits should be familiar for all who have noticed that writing software to mimic something typically enhances one's understanding of the subject itself; indeed, machine ethics research plausibly could become a useful "laboratory" for clarifying ethical theories within moral philosophy. Creation of ethical robots was, of course, a concern of the science-fiction author, Isaac Asimov. In fact, assessments of his celebrated "Three Laws of Robotics" by researchers in robotics and AI were reported during another of the workshop's presentations. Widespread awareness of the laws notwithstanding, these assessments predominantly judged them unsuitable for actual implementation, citing problems such as ambiguity and logical ordering in Asimov's formulations. Application of general prescriptions against harming humans, for instance, can—as one member of a leading robotics institute observed—be difficult to represent in software. Compounding such problems of representation, another workshop presenter added that, according to the ethical theory of theologian Paul Tillich, guidance for application of moral "rules" depends critically upon experience of unconditional love (agape)! On the other hand, the presenter also noted that apparent human dispositions to assign moral status to humanlike robots might produce corresponding challenges for communities endorsing Tillich's ethics. The theologian argued, for example, that moral motivation involved provision of divine grace, which could entail some problematic conclusions within such communities regarding availability of such grace to machines.

Human implications of HRI involving the concept of personhood arose in a presentation that viewed current robots as artifacts possessing instrumental value—but lacking attributes of sentience and freedom needed for full moral status. Under these conditions, the presenter urged, apparent human dispositions to bond with humanlike robots during HRI could result in deleterious "misplaced personal relationships."

The workshop's first afternoon presentation revealed interesting relations between modeling of intention in HRI research and results from neuroscientific study of human intention. Noting experimental evidence that neurological preparation for a human action temporally precedes conscious experience of "intention-in-action," the presenter developed an alternative to traditional symbolic belief-desire-intention models, treating each intention as an ongoing process rather than a static point in a domain of intentions. Implications of this innovation for common folk-psychological understandings of human intention could be significant.

Privacy of personal information often is an issue that people charge with strong feelings. Useful HRI with future robot butlers can be expected to require these artifacts to accumulate substantial amounts of personal information concerning the individuals whom they serve. Accordingly, another of our workshop presentations reported exploratory study of HRI scenarios revealing people's attitudes toward treatment of such information.

Again, medical and health-care applications of robotics are known to be rich, with HRI bearing human implications for patients as well as health-care personnel. An afternoon presentation focusing upon the latter category reported survey results strongly recommending, among other things, that user-centered design be more intensively practiced during development of artifacts intended to assist personnel in this area.

A concluding presentation reported investigation of the complex HRI phenomenon of "mutual adaptation," linking human and robot during nonverbal (gesture) instruction of the robot. For this study, a set of experiments was conducted using Wizard of Oz (WOZ) methods, allowing a remote operator to control the robot involved. The papers from the workshop were published as AAAI Technical Report WS-07-04 and are available from AAAI Press.

—Ted Metzler and Lundy Lewis

Intelligent Techniques for Web Personalization

The web today is an opulent smorgasbord of applications in which users interact with companies, the government, and a wide variety of information providers. However, the potential of the web is hampered by the enormity of the content available and the diverse expectations of its user base. These challenges, in turn, have driven the increasing need for more intelligent, personalized, and adaptive web services or applications, such as e-commerce recommender systems. Businesses have come to realize the potential of these personalized and adaptive systems to increase sales and retain customers. Likewise, web users have come to rely on such systems to help them more efficiently find items of interest in large information spaces.

Web personalization and recommender systems are closely related and partially overlapping areas of research. To bring together researchers from both areas and foster the exchange of ideas between the involved communities, the two AAAI-2007 workshops, the Fifth Workshop on Intelligent Techniques for Web Personalization (ITWP'07) and the Workshop on Recommender Systems in E-Commerce, joined forces to address issues and challenges in the design, implementation, deployment, and evaluation of web personalization and recommendation solutions, both from a research as well as from a practical perspective. The technical papers presented at the workshop were selected based on a rig-

orous review by the internationally renowned program committees of the workshops and covered topics like customer behavior modeling, preference elicitation, scalable and effective recommendation algorithms, personalized search and information access, data and web mining for personalization, trust and security in recommender systems, as well as aspects of the evaluation of such systems.

Several papers presented at the workshop were devoted to the problem of improving the accuracy and effectiveness of recommender systems. They noted that improvements could be achieved by taking additional (context) information into account, by applying novel rating measures when generating recommendations for users, or by considering semantic knowledge about the domain from an available ontology. Several contributions focused on practical applications of recommender systems in areas such as health services, tourism, and legal domains. Another important topic examined by some papers was the security and robustness of different types of recommender systems in the face of shilling or profile injection attacks.

Search personalization and techniques for the automated construction and maintenance of user profiles were the prominent areas of interest in the web personalization tracks. A newly proposed approach considered clustering user sessions aimed at reformulating search engine queries according to user profiles. Another contribution proposed a novel approach: to incrementally learn ontological user profiles to personalize search results. Finally, search personalization was considered in the context of the social web (web 2.0) in a paper proposing an approach for personalizing image search results on Flickr.

In addition to the technical papers, the workshop also featured presentation of late-breaking ideas during a poster session. This informal session particularly helped stimulate further discussion among the more than 40 attendees of the workshop. The workshop ended with an open discussion of further developments and open research challenges in the field, with a particular focus on practical aspects of web personalization and recommendation solutions of the future.

The proceedings of the workshop, including 13 long papers and 6 short papers authored by researchers from 11 different countries, are available as Technical Report WS-07-08 from AAAI Press.

—Bamshad Mobasher, Sarabjot Singh Anand, Alfred Kobsa, and Dietmar Jannach

Plan, Activity, and Intent Recognition

The Plan Activity and Intent Recognition (PAIR-07) workshop was the latest in a series of successful workshops designed to bring together researchers working in a number of different communities all related to modeling and recognizing the actions, behavior, and goals of human and synthetic agents. The work in this area is done under a number of different research headings, and this has resulted in fragmentation in the field. We felt that a wider sharing of results would help researchers working in this area. Therefore, over the last few years, we have organized the modeling others from observations (MOO) series of workshops (at AAMAS, IJCAI, and AAAI) to focus on these issues.

As successful as these workshops have been, at the MOO-06 workshop the organizing committee felt that changing the name to focus on recognition and deemphasize the modeling aspects might attract a wider set of researchers. This change in focus seems to have had the desired effect as was witnessed in the diversity of presented research and application areas. PAIR-07 had a wide range of research areas including traditional plan recognition, learning by demonstration, preference identification, hostile activity recognition, and even building mental models of other agents. All of these applications made use of plan, activity, or intent recognition technologies but made use of a wide range of theoretical approaches.

However, even given this diversity of application areas and approaches, there were common themes that emerged. As had occurred at previous workshops, we saw an emphasis on the use of probabilistic reasoning for recognition. This covered use of traditional methods, such as Markov random fields, and decision-theoretic approaches to hybrid logical and probabilistic methods. The application of these technologies for plan activity and intent recognition is an exciting extension beyond the early strictly logical work in this area and continues to present the possibility of significantly extending the capabilities of deployed systems.

One of the new themes emerging at PAIR-07 was the use of more real world data. While some of the work reported in earlier workshops was based on real-world data, PAIR-07 saw in increase in systems that were evaluated on such real-world data extracted from application domains like assistive systems for the elderly and eBay auctions. This is a very exciting development since it both indicates the maturity of the technology and demonstrates a need within applied systems for the research being developed. It is also clearly driving many researchers to extend the existing approaches and formalisms to deal with real-world phenomena.

A number of attendees were very excited about continuing these discussions and are interested in looking for more ways to encourage still wider participation in future workshops. We are very optimistic about the advances reported on at PAIR-07 and believe it has helped bring the community together to more widely spread these significant research results. Such interactions are critical for our community to thrive and grow.

The papers from the workshop were published as AAAI Technical Report WS-07-09 and are available from AAAI Press.

—Christopher Geib and David Pynadath

Preference Handling for Artificial Intelligence

Most problems studied in artificial intelligence involve some form of choice. For example, a robot has to choose among alternative plans to reach given goals, a web-based recommender system should choose a con-

figuration that pleases the user, and an automatic translation system has to struggle with the multiple meanings of words. All these problems may have huge spaces of possible decisions that significantly differ in general criteria, such as cost, quality, simplicity, as well as domain-specific criteria. Preferences are a convenient way to compare the options a priori and then to use them to make best choices in a multitude of problems with different decision spaces.

Preferences are thus essential for making choices in a rational (and intelligent) way. Preference models have been necessary in many fields of AI, including multiagent systems, combinatorial auctions, knowledge representation and reasoning, planning, diagnosis, and design. Moreover, preference modeling is central to decision theory, social choice, and game theory, which, more and more, are cross-fertilizing with AI. AI brings new problems to these classic fields and often needs new forms of preference handling beyond classic utility-based models such as graphical and logical preference representations, new forms that can directly be used in preference-based problem-solving algorithms.

Preference handling has become an intense area of research in AI. The AAAI-07 workshop continued a series of successful workshops on preference handling at AI conferences (AAAI-02, a Dagstuhl-seminar in 2004, IJCAI-05, and ECAI '06). The one-day workshop had a dense program of 14 presentations and a discussion about the role of preference for AI. Furthermore, AAAI-07 included a tutorial, an invited talk, and many technical papers about preferences.

The workshop papers applied preferences to a variety of AI problems such as e-commerce and combinatorial auctions, intelligent assistants, winner determination in majority voting, game theory, search for solving combinatorial problems, configuration, meeting scheduling, peer-to-peer query answering systems, geographic map generation, and conference paper assignment. The last two problems unveiled interesting preference structures. The geographic map generation chooses one among several satellite images for each position in a grid while imposing preferences on adjacent images in addition to preferences on the choice of a single image. Assigning conference submissions to program committee members may bear a multitude of criteria such as "minimize the number of papers that are reviewed from reviewers from the same institute" in addition to classic bid maximization. Both cases show that preference aggregation can be nontrivial in practice.

Another track of papers addressed important questions concerning the elicitation, learning, and representation of preferences, in particular when those preferences are formulated over a combinatorial criteria space and when the different criteria are not preferentially independent. The workshop explored different forms of limited preferential independence that allow a compact, usually graphical form of those sophisticated, but often realistic, user preferences. Several open questions were discussed during the workshop, such as giving a precise meaning for preferences that reverse more general preferences. For example, a user may generally prefer cheap cameras to expensive ones and prefer light cameras to heavy ones but then finally buy an expensive heavy camera since it has other advantages.

The workshop finished with a discussion about the role of preferences for artificial intelligence. The audience remarked that preferences are necessary for robotics and intelligent agents who need to make decisions. Although the community of AI researchers interested in preference handling is growing from one year to another, general awareness of the potential and importance of preference handling for AI is still missing. The participants therefore discussed new ways for disseminating results about preference handling to the AI researchers and practitioners and to stimulate more exchange with classic AI applications such as robotics.

As the workshop focused on questions from AI, it was complementary to the multidisciplinary workshop on advances in preference handling that was held at the very large data base conference (VLDB) in 2007.

The papers from the workshop were published as AAAI Technical Report WS-07-10 and are available from AAAI Press.

—Judy Goldsmith, Ulrich Junker, Jerome Lang, and Jon Doyle

Semantic e-Science

As a synthesis of knowledge and web technologies, the semantic web has been gaining tremendous attention world-wide. One noteworthy movement is the increasing recognition of this new technology in typical scientific areas such as life science, earth science, and social science. In general, the merits of the semantic web with respect to scientific research are exhibited in different perspectives encompassing collective knowledge acquisition, semantic data mashup and integration, scientific workflow management and composition, integrative knowledge discovery and data mining, logic-based hypothesis checking, and so forth.

However, semantic web researchers have, to date, largely focused on the formal aspects of logic languages or general-purpose semantic application development, with insufficient consideration of requirements from specific scientific areas. On the other hand, general science researchers grow ever more dependent on the web, but they have no coherent agenda for exploring the emerging trends on the semantic technology. It urgently requires the development of a multidisciplinary field to foster the growth and development of e-science applications based on the semantic technologies and related knowledge-based approaches. Advances in e-science infrastructure and e-science applications based on the semantic technologies and related knowledge-based approaches call for increased interaction among these disparate communities. The AAAI semantic e-science workshop provided an interdisciplinary forum for researchers coming from the semantic web community and general science communities.

For 2007, we received 25 submissions. Eight papers were accepted as regular papers, and four as short papers. The covered topics included

semantic-based data integration, ontology-based service composition, data privacy, and trust management in e-science applications, data mashup, data policy, data provenance, and manipulation trace, semantic graph mining, and knowledge discovery. The applications encompassed biomedicine, geology, citizen science, earth science, and social science.

Two keynote speakers were invited and presented their viewpoints from different perspectives. Deborah L. McGuinness stated that it was crucial to support explanations of provenance, information manipulation trace, and trust using an interoperable, transparent, and user-friendly knowledge provenance infrastructure, as systems become more and more complex and decentralized. Yolanda Gil presented her work on managing complex scientific workflows in support of large-scale distributed data mining and scientific analysis.

There has been increasing interest in the application of semantic web technologies to numerous scientific domains. We believe that semantic technologies will play an increasingly important role and exert greater influence in accelerating and facilitating scientific research cycles.

The papers from the workshop were published as AAAI Technical Report WS-07-11 and are available from AAAI Press.

—Huajun Chen, Yimin Wang, and Rudi Studer

Spatial and Temporal Reasoning

The AAAI-2007 workshop on spatial and temporal reasoning continued a series of workshops that started 14 years ago at IJCAI-93 in Chambery, France. The common goal of these workshops has been to bring together related communities of researchers with an interest in the study of representing and reasoning about either space or time—or both.

Recent years have witnessed remarkable advances in some of the long-standing problems in the field of spatial and temporal reasoning (for instance, new results about tractability for spatial calculi, explicit construction of models, characterization of important subclasses of relations), as well as in the development of new areas (the appearance of new integrated spatio-temporal calculi is one example, as well as the development of multidimensional spatial calculi). Likewise, proposals have been made to remedy some of the weak points of the symbolic approach by introducing fuzzy versions of classical calculi or importing nonmonotonic techniques for dealing with incomplete information. Despite all these efforts, there is still a lack in a deeper understanding of the foundations of the field, which might be the reason that it has not found as much enthusiasm among information technology practitioners as it should have had. The workshop was aimed at gaining such a fundamental understanding, focusing on three causes for the current situation: (1) fundamental—no existing generalized understanding across different domains of space and time; (2) methodological—no formal general-purpose methodology has been developed across different spatio-temporal calculi studied, making it difficult to compare and contrast between these disparate calculi; (3) strategic—the lack of a critical mass of application fields for each individual spatial or temporal calculus owing to the above two reasons.

The workshop was intended both as a forum for discussion, exchange of points of view, and assessment of results and methods and as a source of dissemination and promotion of the newest advances in the area. Participants came from around the world, with a slight dominance of North American participants. The presentations were almost equally split between the area of spatial reasoning on the one hand and the area of temporal reasoning on the other. In general, the contributions were of high quality and led to interesting and inspiring discussions.

Although progress was made in various areas of spatial and temporal reasoning, it became obvious that there are still a significant number of open problems. While debating interesting theoretical challenges, one question repeatedly surfaced during the discussions: What is the value of all the theories in regard to practical applications in AI and beyond? Diverging opinions emerged regarding this important issue, covering the whole spectrum between the two extremes: on the one hand, the field will survive only if the theoretical challenges are addressed, and on the other hand, practitioners will never need the rich formalisms that are currently developed. Although no consensus was reached regarding this question, it became clear that further work is needed in the field. The intention therefore is to organize follow-up workshops at other international, American, or European AI conferences.

The papers from the workshop were published as AAAI Technical Report WS-07-12 and are available from AAAI Press.

—Hans W. Guesgen, Gerard Ligozat, Jochen Renz, and Rita V. Rodriguez

Trading Agent Design and Analysis

Trading agents have become a prominent application area in AI because of their potential benefits in electronic commerce and because they present a stiff challenge to models of rational decision making. This workshop focused on the design and analysis of agents that must operate in open, uncertain, competitive economic environments. Submissions were invited to focus on trading agent architectures, decision-making algorithms, empirical evaluations of agent strategies in negotiation scenarios, and game-theoretic analyses.

The workshop was held in conjunction with the finals of the 2007 Trading Agent Competition. Two game scenarios and two challenge events attracted 39 entries. The supply-chain management scenario placed six agents in the role of a PC manufacturer. Each agent had to procure raw materials and sell finished goods in competitive markets while managing inventory and production facilities. The procurement challenge was a side competition that allowed agents to balance risk and cost in the procurement market by providing both long-term and short-term contracts. The

prediction challenge was a second side competition to test price-prediction capabilities of competing agents in both procurement and sales markets.

Rather than cast the competing agents as traders, the CAT scenario placed agents in the role of competing exchanges. This competition was motivated by the rise of independent for-profit stock and commodity exchanges that compete for the attention of traders. CAT agents competed by defining rules for matching buyers and sellers and by setting commission fees for their services. Profitability was the ultimate measure of performance in both the supply chain and CAT scenarios.

The day featured six extended paper presentations, detailed discussions of the game scenarios and the performance of the competitors, and a panel discussion on the future of the competition and its effectiveness in supporting research. Three of the papers focused on various aspects of performance analysis for the supply chain management (SCM) scenario; one included work that extended techniques in empirical game theory, one proposed a "repeated simulation" method to control variability, and one applied information gain methods to various aspects of performance analysis. This work is important because the complexity and uncertainty in the game scenario make it difficult to understand why one agent outperforms another in general or in particular market conditions. One paper presented preliminary results on the performance of market-maker agents in the CAT scenario, evaluating the relative performance of various well-known strategies in the competitive simulation environment. This is important because the resulting benchmark data lays the groundwork for future empirical research in this area.

The day ended with a panel and group discussion on the future of trading agent research and supporting competitions. It was felt that the various trading agent competitions have stimulated a rich body of research, and yet there remains much to be learned. The challenge competitions were thought to be quite valuable in focusing attention on individual aspects of agent performance. The procurement challenge was also seen as a dry run for a possible new version of the supply chain scenario, with a richer interaction model between agents and suppliers, allowing a mix of short-term and long-term procurement contracts. In the longer run, there is considerable interest in finding ways to enable human versus agent competition in a scenario similar to the supply chain game. The current simulation design would present a human player with an overwhelming volume of data; the challenge, as it is in the real world, would be to find ways to put the human in control of important decision factors without the necessity to make large numbers of individual pricing, procurement, and resource allocation decisions.

The papers presented in this workshop were published as an AAAI technical report WS-07-13. The chair of this workshop was John Collins (University of Minnesota).

—John Collins and Michael Wellman

Information Integration on the Web

Information-integration techniques enable interaction between users and data sources through a centralized access point and uniform query interfaces that give users the illusion of querying a homogeneous system. Most integration solutions have assumed structured sources with the heterogeneity being introduced by the variety in source schemas and data models. But a large portion of the web consists of pages that contain information presented as unstructured text, such as, for example, blogs, wikis, reviews, and so on. Therefore, integration systems that can match entities or objects across both structured and unstructured sources are the need of the hour. Recent research on web object extraction, record linkage, and named entity recognition has generated some initial solutions; however, many challenges remain in developing such a system.

This workshop, sixth in the IIWeb series, was proposed to bring together researchers looking to effectively integrate information from both structured and unstructured sources of information. The anticipated outcome of the workshop was to assess the state of the art in the area, as well as to identify critical next steps to pursue in this topic.

Accordingly we asked submissions in topics such as automatic wrapper induction, schema matching, web entity extraction and search, record linkage and object consolidation, database and information retrieval (IR) integration, applications, and experiences. The call attracted 24 submissions out of which 12 were selected for full presentation and 6 for short presentation.

Phillip Bernstein from Microsoft Research gave the keynote speech, "A Schema Mapping Infrastructure for Data Integration," at the opening of the workshop. The keynote provided a comprehensive picture of the schema-mapping problem and the state-of-the-art solutions, and it was well received.

The accepted 12 full (20 minutes) and 6 short (10 minutes) presentations were grouped into five sessions. The session topics were novel integration architectures and web applications; online information integration, name disambiguation, and entity resolution; semantic data integration and schema matching; and web information extraction.

For this workshop, each session was followed by a short panel discussion. The session chairs were the moderators. We had 15 minutes for people to ask questions with respect to all the presentations in the session and the future research directions of the topic. This idea worked very well, and people were actively participating in discussions.

The workshop was well attended. Although one paper was presented using recorded video because the presenter had trouble obtaining a U.S. visa, the other 17 presentations were made in person. About 30 people attended the workshop, and most of them stayed until the very end. The papers from the workshop were published as AAAI Technical Report WS-07-14 and are available from AAAI Press.

—Ullas Nambiar and Zaiqing Nie