Tutorial Long Description for AAAI-18 conference *– version of January 15, 2018*

Title: "Rulelog: Highly Expressive Semantic Rules with Scalable Deep Reasoning Networks"

Presenters: Benjamin Grosof (Accenture), Michael Kifer (Stony Brook University), Paul Fodor (Stony Brook University), and Janine Bloomfield (Coherent Knowledge); benjamin.n.grosof@accenture.com, kifer@cs.stonybrook.edu, pfodor@cs.stonybrook.edu, janine.bloomfield@coherentknowledge.com (see also CVs/contact below).

Duration: half-day (3.5 hours, excluding 30-minute break).

Subarea within AI / Keywords: Knowledge Representation and Reasoning (KRR), Cognitive Computing, Semantic Rules, Deep Semantics of Natural Language, Knowledge Integration for Machine Learning.

Abstract: In this half-day tutorial, we cover the fundamental concepts, key technologies, emerging applications, recent progress, and outstanding research issues in the area of Rulelog, a leading approach to fully semantic rule-based knowledge representation and reasoning (KRR). Rulelog matches well many of the requirements of cognitive computing. It combines deep logical/probabilistic reasoning tightly with natural language processing (NLP), and complements machine learning (ML). Rulelog interoperates and composes well with graph databases, relational databases, spreadsheets, XML, and expressively simpler rule/ontology systems – and can orchestrate overall hybrid KRR. Developed over the last 25 years, Rulelog is *much* more feature-full than the previous state-of-the-art practical KRR approaches, yet is computationally affordable. It is fully semantic and has capable efficient implementations that leverage methods from logic programming and databases, including dependency-aware smart caching and a dynamic compilation stack architecture.

Additional Details (continuing from the Abstract): Rulelog extends Datalog (database logic) with general classical-logic-like formulas – including existentials and disjunctions – and strong capabilities for meta knowledge and reasoning, including higher-order syntax, flexible defeasibility and probabilistic uncertainty, and restraint bounded rationality that ensures worst-case polynomial time for query answering. A large subset of Rulelog is in draft as an industry standard. An exciting research frontier is that Rulelog can combine closely with NLP to both interpret and generate English, including potentially for conversational NL interaction.

The most complete system today for Rulelog is Ergo from Coherent Knowledge, which is available free for academic/research use. A subset of Rulelog is also implemented in an open-source Flora-2 system and an earlier SILK system from Vulcan. Using Ergo, we will illustrate Rulelog's applications in deep reasoning and representing complex knowledge – such as policies, regulations/contracts, science, and terminology mappings – across a wide range of tasks and domains in business, government, and academe. Examples include: legal/policy compliance, e.g., in financial services; financial reporting/accounting; health care treatment guidance and insurance; education/tutoring; security/confidentiality policies; and e-commerce marketing.

Prerequisite knowledge: The tutorial will cater to those first learning about declarative logic programs and Rulelog, as well as those who already have some background in them. It will assume only background knowledge of the basics of logical knowledge representation and reasoning: familiarity with the concepts of first order logic and relational databases. Most AI researchers (including students) who attend AAAI have this prerequisite background. Some degree of familiarity with the concepts in one or more of the following will be helpful but is not required: graph databases (RDF/OWL and

SPARQL), declarative logic programs, ontologies, machine learning, Prolog, production rules, Bayesian probabilistic reasoning, and natural language processing.

Goal of the tutorial: The target audience is those who are interested in learning more about knowledge representation and reasoning (KRR) as a core area in AI, and about how it relates to AI's applications as well as conceptual architecture. Together with machine learning (ML) and NLP, KRR forms the tripod basis for the core of AI and cognitive computing. Compared to previous tutorials about Rulelog, this tutorial will cover more about Rulelog's relationships to NLP, ML, and probabilistic uncertainty, including how Rulelog provides a basis for deep semantics of natural language and how Rulelog provides deep reasoning networks that can work in dialectical tandem with ML. The audience will walk away with an understanding of Rulelog's key innovative logical and inferencing concepts, its broad applicability, its overall advantages and limitations, a sample of some specific application areas, and its open research topics.

Outline: (Note: Examples, and discussion, are sprinkled throughout)

- Introduction: logical knowledge representation; practical logic.
- Rulelog features and software: strong meta expressiveness; reasoning methods; scalability; integration points with databases, ontologies, NLP, and ML.
    - Concepts and foundations, in depth: Datalog, well-founded semantics, "tabling" algorithms, rule-based mapping between text phrases and logic syntax, higher-order syntax (HiLog), defeasibility via argumentation rules, general quantified formulas (omniformity), probabilistic and weighted uncertainty, restraint bounded rationality, frame syntax, rule identifiers and provenance, external querying and virtual data stores, updating and reactiveness, integrity constraints and alarms, explanations, rule-based ontologies and ontology mapping.
- Textual Rulelog: concepts and methods for NL understanding and NL generation in and with Rulelog KRR, especially for semi-automatic knowledge authoring, automatic explanation, and automatic terminology mapping. Use of higher-order syntax, NL dependency parsing, and NL co-reference analysis.
- Combination of Rulelog KRR with ML: concepts and methods, including explainable ML, knowledge accumulation, knowledge integration of ML results, supplying derived data to ML, driving ML with subgoal queries, and relationships to Bayesian probabilistic and triangular-norm based non-Bayesian uncertainty.
- Case study demos, with feature tour: financial regulatory compliance; health care treatment guidance.
- Applications overall: Horizontally: policy-based decisions, info Integration, analytics, human-computer interaction (HCI), search, business intelligence, risk management. Vertically: e-commerce & marketing, financial services, personalized e-learning, security & defense, biomedical, insurance, Internet of Things (IoT), social media sharing policies.
- Open research topics in: authoring rules starting from NL; distributed reasoning; optimization of uncertainty reasoning; equality; aggregates; integration with ASP, "constraint solving", and classical logic; hypotheticals; abduction; integration with ML.

More about what makes the topic innovative and important: Rulelog combines several fundamental advances in KRR for complex knowledge (including semantics, theory, and algorithms). It has practical efficient implementations and emerging industrial and academic tools and applications. Greatly

extending Datalog, Rulelog combines concepts and features that until its advent were not found *together* in any practical logical KRR, including versatile defeasibility machinery, probabilistic uncertainty, higher-order general formulas with existentials and disjunctions, and fully semantic bounded rationality enabling polynomial-time scaling. It is perhaps the most flexible and widely-applicable logical KRR available today for deep reasoning and representing complex knowledge – such as regulations/policies, science, and information integration mappings. It has increasingly close relationships to natural language processing and machine learning, the combination with which constitutes the core of cognitive computing. There are many exciting open research opportunities both in extending and optimizing the Rulelog KRR itself and in exploring its applications, which are increasingly broad. Both (1) ML, including knowledge extraction from NL, and (2) graph databases/knowledge, including RDF and SPARQL, have each become hot recently in industry as well as research. But both need to be combined with more complex human-authored knowledge and deeper reasoning in order to deliver business/social value more effectively. Techniques for developing such human-authored complex knowledge, including starting from text, have progressed substantially. A large subset of Rulelog is in draft as an industry standard to be submitted to RuleML and W3C as a dialect of Rule Interchange Format (RIF).

Presenters' Bios – including background in the tutorial area:

Benjamin Grosof (lead presenter; http://benjamingrosof.com) is a Principal Director and Research Fellow in Artificial Intelligence at Accenture. He is an industry leader in AI knowledge representation, reasoning, and acquisition. He has pioneered semantic technology and industry standards for rules combined with ontologies, their acquisition from natural language, and their applications in finance, e-commerce, policies (including contracts, regulations, and security), and e-learning. He co-founded Coherent Knowledge, a software-centric startup that is commercializing a major research breakthrough in AI logical/probabilistic knowledge representation and reasoning (Rulelog) combined with natural language processing, and was its CTO and CEO (2013-2017). He also was president of the expert consulting firm Benjamin Grosof & Associates, founded while he was at MIT (2000-2017). Previously he was a senior research program manager at Vulcan Inc. (2007-2013), where he conceived and led the Advanced Research third of the predecessor of the Allen Institute for AI. Before Vulcan, he was an IT professor at MIT Sloan (2000-2007), where he was also a DARPA PI, and a senior software scientist at IBM Research (1988-2000). He co-founded the influential RuleML industry standards design effort and prototyped it in SweetRules, the main bases for the W3C Rule Interchange Format (RIF) standard. He co-founded the International Conference on Rules and Rule Markup Languages for the semantic web (which since became the RR and RuleML conferences and then the International Joint Conference on Rules and Reasoning). He led the invention of several fundamental technical advances in knowledge representation, including courteous defeasibility (exception-case rules), restraint bounded rationality (scalability in complex reasoning), and rule-based description logic ontologies (the basis for W3C's OWL RL standard). He also has extensive experience in user interaction design, and in combining logical methods with machine learning and probabilistic reasoning uncertainty. His background includes 5 major industry software releases, 2 years in previous software startups, a Stanford PhD in AI, a Harvard BA in applied math, 3 patents, and over 60 refereed publications.

Grosof has given numerous invited talks about knowledge representation, reasoning, and acquisition, including with semantic rules, and developed several MIT courses with substantial focus on those. He presented – with coauthors, including usually Michael Kifer since 2009 – related tutorials on reasoning with complex knowledge at the AAAI Conference on Artificial Intelligence (2013, 2017), International

Joint Conference on Artificial Intelligence (2001, 2016), the International Conference on Knowledge Capture (K-CAP 2015), International Semantic Web Conference (2004, 2005, 2006, 2009, 2010, 2012), the WWW conference (2006, 2009), the International Joint Conference on Rules and Reasoning (2017), the International Web Rule Symposium joint with the Reasoning Web Summer School (2015), and the ACM Conference on E-Commerce (2004).

Michael Kifer (http://www.cs.stonybrook.edu/~kifer) is a Professor with the Department of Computer Science, Stony Brook University, USA. He received his Ph.D. in Computer Science in 1984 from the Hebrew University of Jerusalem, Israel, and the M.S. degree in Mathematics in 1976 from Moscow State University, Russia.

Kifer is a co-founder and CTO of Coherent Knowledge, a startup on semantic technology, and since 2012 he has been serving as the President of the Rules and Reasoning Association (RRA). His interests include Web information systems, knowledge representation, and database systems. He has published four text books and numerous articles in these areas. In particular, he co-invented F-logic, HiLog, and Transaction Logic, which are among the most widely cited works in Computer Science and, especially, in Semantic Web research. Kifer serves on the editorial boards of several computer science journals and chaired a number of conferences. Twice, in 1999 and 2002, he was a recipient of the prestigious ACM-SIGMOD "Test of Time" awards for his works on F-logic and object-oriented database languages. In 2013, Kifer's paper on Transaction Logic Programming was awarded the Association of Logic Programming's “Test of Time” award as the most influential paper of 20 years ago. In 2006, Kifer was Plumer Fellow at Oxford University's St. Anne's College and, in 2008, he received SUNY Chancellor's Award for Excellence in Scholarship. He has taught numerous courses at Stony Brook University since 1984.

Paul Fodor (http://www.cs.stonybrook.edu/~pfodor) is a Research Assistant Professor with the Department of Computer Science, Stony Brook University, USA. He received his Ph.D. in Computer Science in 2011 from the Stony Brook University, New York, preceded by his M.S. degree in 2006 from Stony Brook University, and B.Sc. in Computer Science in 2002 from the Technical University of Cluj-Napoca, Romania.

Dr. Fodor is a co-founder and Senior Engineer of Coherent Knowledge with over 10 years’ experience in databases research, natural language processing, artificial intelligence and stream processing systems. His work on declarative rule languages and logic used as a specification language and implementation framework for knowledge bases was applied in areas ranging from natural language processing to complex event processing and semantic Web technologies. Through his research, Dr. Fodor has contributed to several large software projects: the IBM Watson natural language processing system for the Jeopardy! Challenge with human champions, the OpenRuleBench suite of benchmarks for analyzing the performance and scalability of rule systems for the semantic Web, the ETALIS declarative complex event processing and stream reasoning system, and the SILK Semantic Inferencing on Large Knowledge. Dr. Fodor was Principal Investigator (PI), Co-PI and contractor for projects funded by both public governmental sources and private companies, such as, PI for the SILK project funded by Vulcan Inc. to develop intelligent textbooks, contractor for the IBM Watson project, contractor for XSB Inc. for the DARPA Component, Context, and Manufacturing Model Library (C2M2L-1) using XSB Prolog, and PI for the Stony Brook University Hospital's Lung Cancer Evaluation Center management program. He has taught numerous courses at Stony Brook University since 2011.

Janine Bloomfield (http://coherentknowledge.com; http://www.linkedin.com/in/janinebloomfield) is a co-founder and Chief Operating Officer of Coherent Knowledge, with over 4 years experience there in developing proof-of-concept applications and tutorial materials about Rulelog knowledge representation and reasoning. She received her Ph.D. in ecosystems ecology in 1993 from Yale University, preceded by her M.S. degree in biology from Stanford University. Dr. Bloomfield has experience in knowledge engineering; marketing, writing, and website development, for research and non-research audiences; education, including teaching and curriculum development; data analysis and statistics; and scientific research in life sciences. She has published at the International Conference on Web Rules and Reasoning (RuleML), and presented on Rulelog concepts and applications at the Smart Data (previously named Semantic Technology) industry conference. Previously in her career, she was a senior scientist, on global climate change, at Environmental Defense Fund, one of the most influential USA environmental non-profits. There, she did science communications at national and international level, including: national media; state and federal policy advising; and science curriculum development. She authored a series of influential white papers, op-eds, and web material. Before that, she was a researcher on acid rain at US Forest Service. She also has experience in small-business administration and as a mindfulness educator for students, teachers, and parents.