Foreword

Large amounts of data increasingly becoming available and described using real-life ontologies represented in Semantic Web languages, recently opened up the possibility for interesting real-world data mining applications on the Semantic Web. However, exploiting this global resource of data requires new kinds of approaches for data mining and data analysis that would be able to deal at the same time with its scale and with the complexity, expressiveness, and heterogeneity of the representation languages, leverage on availability of ontologies and explicit semantics of the resources, and account for novel assumptions (e.g., "open world") that underlie reasoning services within the Semantic Web.

The workshop tried to address the above issues, in particular focusing on the problems of how machine learning techniques, such as statistical learning methods and inductive forms of reasoning, can work directly on the richly structured Semantic Web data and exploit the Semantic Web technologies, what is the value added of machine learning methods for the Semantic Web, and what are the challenges for developers of machine learning techniques for the Semantic Web data, for example in the area of ontology mining.

The workshop was meant to bring together researchers and practitioners interested in the interdisciplinary research on the intersection of the Semantic Web with Knowledge Discovery and Machine Learning, and provide a meeting point for the related communities to stimulate collaboration and enable cross-fertilization of ideas.

Specifically, the review phase produced a selection of 5 full papers, 1 position paper, and 2 late breaking news abstracts. IRMLeS 2010 program was further enriched by two invited talks from prominent researchers. Dr Melanie Hilario presented in her talk an ongoing research on optimizing the knowledge discovery process through the semantic meta-mining, involving background ontology representing the domain of data mining. Professor Steffen Staab demonstrated in his talk how the enrichment of Web 2.0 data by automatically discovered semantic relationships may improve the user experience. The workshop was also successful in terms of registrations and attendance.

The topics covered by IRMLeS 2010 included: ontology learning, and semantic tagging to expose the semantics of unstructured or semi-structured data as text, or Web 2.0 tags; management, and retrieval of Semantic Web resources, e.g. RDF data; probabilistic approaches; similarity measures for ontological data; inductive reasoning with ontologies; finally using ontologies, and other formal representations as background knowledge to steer whole knowledge discovery process.

In the final wrap-up discussion, a number of open problems and promising directions were elicited. Similarly as the last year, the topic of integration of induction and deduction has been recognized as essential for the Semantic Web, to deal with real, noisy data. Related to this topic, the topics of probabilistic approaches, and uncertain inference over semantic resources were discussed. The need for new metrics for evaluating the output of machine learning methods in the Semantic Web setting was recognized, especially in the context of the open

world assumption. The novel topic of semantic data mining also gained attention during discussion, and a call for gathering the community of people working on ontologies/another KR formats for representing data mining domain has been issued. Some other new tasks have also been identified as an interesting future direction of research on machine learning for the Semantic Web that include: ontology repair, and instance matching (especially in the context of a lack of unique name assumption on the Semantic Web).

Given such open issues and the success of the two first editions, we plan to organize next edition in the near future.

Heraklion, May 31, 2010

Claudia d'Amato
Nicola Fanizzi
Marko Grobelnik
Agnieszka Ławrynowicz
Vojtěch Svátek

Program Committee Members

- Sarabjot S. Anand – University of Warwick
- Bettina Berendt – Katholieke Universiteit Leuven
- Abraham Bernstein – University of Zurich
- Floriana Esposito – University of Bari
- Mohand-Said Hacid – University Lyon 1
- Melanie Hilario – University of Geneva
- Andreas Hotho – University of Kassel
- Jose Iria – IBM Research, Zurich
- Ross D. King – University of Aberystwyth
- Jens Lehmann – University of Leipzig
- Francesca A. Lisi – University of Bari
- Thomas Lukasiewicz – Oxford University
- Matthias Nickles – University of Bath
- Sebastian Rudolph – University of Karlsruhe
- Jetendr Shamdasani – University of the West of England
- Steffen Staab – University of Koblenz-Landau
- Umberto Straccia – ISTI-CNR, Pisa
- Volker Tresp – Siemens, Munich

Additional Reviewers

Jörg-Uwe Kietz – University of Zurich

Workshop Homepage

`http://irmles.di.uniba.it/2010/`