



AVVISO DI SEMINARI

Dr. Sašo Džeroski

Jozef Stefan Institute, Department of Knowledge Technologies, Lubiana (Slovenia)

Venerdì 16 aprile 2010 ore 11.15, Aula Gödel, II piano, Dipartimento di Informatica

TOWARDS A GENERAL FRAMEWORK FOR DATA MINING

In this talk, we will address the ambitious task of formulating a general framework for data mining. We discuss the requirements that such a framework should fulfill: It should elegantly handle different types of data, different data mining tasks, and different types of patterns/models. We also discuss data mining languages and what they should support: this includes the design and implementation of data mining algorithms, as well as their composition into nontrivial multistep knowledge discovery scenarios relevant for practical application. We proceed by laying out some basic concepts, starting with (structured) data and generalizations (e.g., patterns and models) and continuing with data mining tasks and basic components of data mining algorithms (i.e., refinement operators, distances, features and kernels). We next discuss how to use these concepts to formulate constraint-based data mining tasks and design generic data mining algorithms. We finally discuss how these components would fit into the overall framework and in particular into a language for data mining and knowledge discovery.

Sašo Džeroski is a scientific councillor at the Jozef Stefan Institute and associate professor at the Jozef Stefan International Postgraduate School, both in Ljubljana, Slovenia. His work is mainly in the area of machine learning and data mining (constraint-based data mining, data mining query languages, inductive logic programming, relational data mining, equation discovery and inductive databases) and their applications (mainly in environmental and life sciences). He coordinated the EU FP6 IST FET project IQ: Inductive Queries for Mining Patterns and Models, where he developed a number of methods for constraint-based learning of models for structured prediction, as well as a general framework and an ontology for data mining. He has been actively involved in many other EU-funded research projects (ESPRIT III project ILP, ESPRIT IV projects ILP2 and METAL, FP5 projects ECOGEN and cInQ, FP6 projects SIGMEA, IQ and EETPipeline, FP7 project PHAGOSYS), as well as the EU-funded Networks of Excellence in ILP (Inductive Logic Programming) ILPnet and ILPnet2, the latter of which he coordinated. He is co-author/co-editor of ten books/volumes, the latest two being »Computational Discovery of Scientific Knowledge« and »Knowledge Discovery in Inductive Databases« (Springer, Berlin 2007).

Seminario per: 1) Dottorato in Informatica; 2) Progetto formativo PS_092 (DIPIS).