Extracting big data from the web: technology, research, and business

Georg Gottlob
University of Oxford & TU Wien

Host: Vladimir Kolmogorov

Do you need to rent a new apartment fulfilling certain requirements? Or would you just like to find a restaurant in your area that serves pasta al pesto as today’s special? In either case, you would most likely start a web search, but keyword search as provided by current search engines is not really appropriate, as it does not allow you to pose complex queries. Solving this problem, at least for certain verticals such as real estate, used cars, restaurants, requires the extraction of massive data from heterogeneously structured websites and the storage of the data into a database having a uniform schema.

In this talk I will report about my 15 years long venture into Web data extraction. In particular, I will discuss the Lixto project we carried out at TU Wien, and the DIADEM ERC project we recently accomplished at Oxford. I will survey the tools and systems we constructed applications we carried out, and also some research results about the logical and theoretical foundations of Web data extraction we achieved. In addition, I will report about two start-ups we spun out.

Wednesday, November 25, 2015, 4:00pm
Raiffeisen Lecture Hall

This invitation is valid as a ticket for the IST Shuttle from and to Heiligenstadt Station. Please find a schedule of the IST Shuttle on our webpage (note that the IST Shuttle times are highlighted in dark green):
http://ist.ac.at/fileadmin/user_upload/pdfs/IST_shuttle_bus.pdf
The IST Shuttle bus is marked IST Shuttle and has the Institute Logo printed on the side.