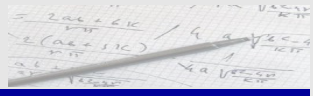
UNIVERSITÀ DELLA CALABRIA

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE



Title: **Extracting Big Data from the Web: Technology, Research and Business**

Speaker: **Georg Gottlob, University of Oxford & TU Wien**

Abstract: 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.

Short Biography: Georg Gottlob is a Professor of Informatics at Oxford University, a Fellow of St John's College, Oxford, and an Adjunct Professor at TU Wien. His interests include data extraction, database theory, graph decomposition techniques, AI, knowledge representation, logic and complexity. Gottlob has received the Wittgenstein Award from the Austrian National Science Fund, is an ACM Fellow, an ECCAI Fellow, a Fellow of the Royal Society, and a member of the Austrian Academy of Sciences, the German National Academy of Sciences, and the Academia Europaea. He chaired the Program Committees of IJCAI 2003 and ACM PODS 2000. He was the main founder of Lixto, a company that provides tools and services for web data extraction. Gottlob was awarded an ERC Advanced Investigator's Grant for the project "DIADEM: Domain-centric Intelligent Automated Data Extraction Methodology".

Date	Time	Room
18/12/2015	16:00	MT10– 30B