



Seminario

Retrieval, Crawling and Fusion of Entity-centric Data on the Web

Dr. Stefan Dietze

L3S Research Center, Hannover, Germany

This talk will provide an overview of approaches to deal with the increasing dynamics and heterogeneity of Web data. In the first part, approaches for focused crawling, linking, profiling and retrieval are discussed, as means to enable discovery and search of entity-centric data in the Web of Linked Data. In the second part, we will turn towards embedded markup, such as Microdata and RDFa, as a novel source of entity-centric knowledge. While markup has seen increasing adoption over the last few years, driven by initiatives such as schema.org and being already adopted by 38% of all Web pages, it constitutes an increasingly important source of entity-centric Web data, where the scale and dynamics are in a similar order of magnitude as the Web (of documents) itself. We will present some case studies and ongoing work on data fusion from markup data which exploit supervised models for aiding tasks such as entity retrieval and knowledge base augmentation. Future directions are concerned with the exploitation of the complementary nature of markup data and traditional knowledge graphs.



UNIVERSITÀ
DEGLI STUDI
DI PALERMO

C.I.C.S.I. Consiglio Interclasse dei
Corsi di Studio in Informatica

d.m.i
matematica e informatica @ unipa



Consiglio Nazionale delle Ricerche
Istituto per le Tecnologie Didattiche

mercoledì 27 settembre 2017
ore 15- Aula 7

Dipartimento di Matematica e
Informatica
Via Archirafi, 34 - Palermo

Evento organizzato nell'ambito dei Seminari di Avviamento al
Lavoro del CICS.I.

Gli studenti partecipanti possono fare richiesta dell'attribuzione
dei CFU previsti per altre attività formative.