



University of Rome “Tor Vergata”

Intelligenza Artificiale 2: Linked Open Data

Introduzione al corso

Armando Stellato <stellato@uniroma2.it>

Organizzazione del corso

Titolo del corso: Intelligenza Artificiale 2

Sottotitolo: Da Open Data a Linked Open Data: un percorso a 5 stelle per dati aperti

Durata del corso: 1 semestre – 9 crediti (6 di corso + 3 di progetto finale)

Obiettivi del corso

- Padronanza del paradigma di gestione e accesso ai dati distribuito nel web.
- Riesamina dei fondamenti di rappresentazione della conoscenza e inferenza logica appresi in altri corsi della laurea triennale e ulteriore affinamento delle conoscenze su questi temi.
- Progettazione di modelli della conoscenza come ontologie, tesauri e lessici per il Web dei Dati.
- Pubblicazione, accesso e esplorazione di dati nel web. Tecniche di coordinamento semantico basato su metadati.

Programma del corso

- Introduzione ai dati aperti. I 5 livelli di qualità dei dati aperti fissati dalla W3C. Licenze sui dati: licenze libere/aperte. Dati, metadati, vocabolari dei dati, ontologie, tesauri: ambiguità terminologiche in uno scenario inerentemente intersettoriale. Il valore economico dei dati. Open Data per il dominio dei Patrimoni Culturali.
- Il Web Semantico: una introduzione. Il Web Semantico e gli Open Data: Linked Open Data e il Web. Modelli dei dati e della conoscenza per il Web.
- Multilinguità e diffusione: aspetti denotazionali, proprietà terminologiche e modelli di lessico avanzato. Interazione lessico-semantica.
- SPARQL: un linguaggio di interrogazione dei dati RDF & assieme un protocollo di accesso ai dati aperti
- Metadati: importanza dei metadata e nel loro ruolo per l'autonomia delle macchine nel web
- LOD Cloud: la "nuvola" dei Linked Open Data nel Web, un bootstrap del Semantic Web, ma anche una contraddizione in essere
- Dataset repositories e portali; Tecnologie per i Linked Open Data: Triple stores. RDF API & Middleware. Ontology Editors, Thesauri Editors, differenti paradigmi e dimensioni d'uso
- ETL in chiave RDF: triplicazione di sorgenti legacy e/o non strutturate.
- Pubblicazione dei dati aperti: standard & best practices