Information about the course on:

# "Semantic Web and Linked Open Data: motivations, solutions, models and technologies. Applications for Text Retrieval"

**The course is held by**: Armando Stellato

**Affiliated to**: "University of Rome Tor Vergata", Rome, Italy

**Title of the course**: Semantic Web and Linked Open Data: motivations, solutions, models and technologies. Applications for Text Retrieval

**Tentative Schedule for the Course**
I would need some feedback on the course attendants, in order to assess the contents to be teached, so that these are adequate for their background knowledge and their expectations on the lectures. This is a schedule of the program, divided among the four days. It is a tentative schedule, subject to feedback from the audience. I foresee some possible overflow from day 2 to day 3 (and from 3 to 4 as a consequence).

*Approx. Time Schedule per day*
9:30-13:00 First Slot
13:00-14:00 Lunch Break
14:00-17:00 Second Slot

*11.10.16*

1. Introduction to Open Data
2. The five-stars quality classification for open data established by W3C
3. Licenses over data: free/open licenses
4. Data, metadata, data vocabularies, ontologies, thesauri: some clarifications and a few terminology
5. The economic value of data
6. Open Data in the Cultural Heritage domain

*12.10.16*

7. The Semantic Web: an introduction
8. The Semantic Web and Open Data: Linked Open Data
9. Models for the Semantic Web
    a. Data Model: RDF
    b. Knowledge Representation Models: OWL, SKOS, SKOS-XL
    c. A few examples of RDF data models:
        i. FOAF: Friend of a Friend
        ii. WGS84: un microvocabolario per la geolocalizzazione
        iii. Dublin Core: metadati o vocabolario general purpose?
    d. Multilingualism: from rdfs:Label, through the terminological properties of SKOS(XL) to the advanced lexical models and ontology-lexicon interfaces of Lemon/OntoLex.
    e. SPARQL: a query language and and a data access protocol over the web for RDF
    f. Metadata:
        i. VoID: vocabulary of Interlinked Datasets
        ii. VOAF: Vocabulary of a Friend
        iii. LIME: Linguistic Metadata

g. LOD Cloud: the "cloud" of Linked Open Data over the Web, a bootstrap of the Semantic Web, and a contradiction at the same time

h. Dataset repositories: datahub.io

*13.10.16*

10. Thesauri Development: the SKOS and SKOSXL models in detail
11. Technologies for Linked Open Data
   a. Triple store: a "DBMS" for RDF triples
   b. A few notable triple stores:
      i. Sesame2 native&in-memory triple stores
      ii. Jena SDB/TDB
      iii. GraphDB
      iv. Virtuoso
      v. AllegroGraph
   c. RDF API & Middleware: RDF dataset access & query modalities. Differences in approaches and technologies
   d. RDF Middleware, the most notable ones:
      i. Sesame
      ii. Jena
   e. Linked Data Platforms
      i. Apache Clerezza
      ii. Apache Marmotta
   f. Ontology Editors, Thesauri Editors, different paradigms and usage scenarios:
      i. Protege
      ii. Web Protege
      iii. TopBraid Composer
      iv. Semantic Turkey
      v. Pool Party
      vi. VocBench

*14.10.16*

12. Triplification of legacy and/or unstructured information sources: acquisition, analysis, "cleaning"
   a. Information Source types
   b. Issues with data triplification
   c. Solutions and approaches
   d. Available technologies
13. Open Data publication: standard & best practices
14. Experiences of Open Data in the P.A. and in large organizations
   a. FAO: the Agrovoc thesaurus and the "open migration" of data.fao.org
      i. When opening data is just the start: the GACS initiative
   b. Italian Senate of the Republic: the Teseo thesaurus and the data on bills, voting, commissions etc.. in dati.senati.it
   c. Office of Publications of the EU: EuroVoc, its intrinsic value, and its centric role as an hub for data sources of the various UE countries.
15. Designing and developing an ontology for Textbook Research – A first Approach