= Open Data Summer School =

Objective

The aim of the summer school is to introduce to undergrad students (mainly, from the first two years of study) the main technologies, tools and topics that are linked to Open Data and Information Extraction. As these topics are usually studied during the senior year undergrad or even at graduate courses, the participating students will have the change to discover a new domain and to experiment with tools in order to see whether they consider interesting the technologies and topics related to Open Data and Information Extraction.

Dates

The Open Data Summer School is scheduled for 2 weeks (June 16th to June 27th 2014). During the first week the participants will attend a series of lectures presenting the main concepts related to Open Data and Information Extraction. In the second week, they will team up in small groups of 2-3 students in order to develop (with the help and guidance of the organizers) an application that makes use of the concepts and technologies studies during the first week of the summer school.

Schedule

First week: June 16th to June 20th - Lectures and exercises/workshop from 10am-5pm

0. Opening

- June 16th, 11am, room EC101
- Why Open Data? What will you learn?
- Organization details

1. Crash Course in Python

- June 16th
- Python syntax
- Basic packages
- Useful packages for Open Data (data processing, visualization, text processing, http/crawling)
- Workshop: Develop your own Python app

2. Open Data Intro

- June 17th
- What is Open Data?
- Semantic Web and Linked Data
- Data modeling (RDF)
- Workshop: Model data using RDF

3. Open Data Scripting (SPARQL, RDF & Python)

- June 18th
- Querying the Semantic Web
- SPARQL and DBPedia
- Setting up a SPARQL endpoint
- Workshop: Querying DBPedia with Python and SPARQL

4. Technologies for Web Crawling, Indexing and Search

- June 19th
- Web crawling
- Intro to indexing
- Web search basics
- Technologies in Java: Apache Nutch, Lucene, Solr
- Technologies in Python: Scrapy, BeautifulSoup
- Workshop: Crawling and indexing in Python, Generating the web graph, Computation of PageRank

5. Elements of Natural Language Processing

- June 20th
- Basics of Natural Language Processing
- POS Tagging
- Dependency Parsing
- Named Entity Recognition
- Information Extraction
- Workshop: Java/Python for a simple NLP app

6. Building an open data application

- June 23rd-27th
- Decide upon a project idea
- Develop your project
- Project presentation: June 27th, 3pm, EC101

Second week: June 23rd to June 27th - Working on a project, with the help of the tutors, from 10am-5pm

Organizers

- Vlad Posea
- Traian Rebedea
- Costin Chiru
- Mihai Dascalu

Participants

- Roxana Capitanu, undergrad 2nd year
- George Muraru, undergrad 1st year

Developed Projects

Team 1 -

Team 2 -

Team 3 -

Sponsors

Lunch and beverages will be provided with the help of Stagii pe Bune