

C3. WEB DATA COLLECTION METHODS III: NON-REACTIVE DATA COLLECTION

Subject Information:

Code	C3	Plan	2014	ECTS	5
Type of Subject		Year	2015	Semester	second
Knowledge area:	Web data collection methods III: non-reactive data collection				
Department:					
Virtual Platform	Platform:	Studium.usal.es			
	Access URL:	Studium.usal.es			

Data about the instructor-teacher

Teacher	Nicoletta Fornara	Group / s	
Department	Faculty of Communication Science		
Knowledge area	Artificial Intelligence		
Academic Centre and University	Università della Svizzera italiana		
Office Address:	Via G. Buffi 13, 6900 Lugano, Switzerland		
Personal URL:	http://www.people.usi.ch/fornaran/		
E-mail	fornaran@usi.ch	Telephone /Skype User:	+41 (0)58 666 45 13

Remarks (previous requirements, coordination, other, if any)

Objectives and competencies of the subject (basics, general, transversal, specifics)

The objective of this course is to introduce students to the techniques which can be used to collect data on the Internet and on the Web by using non-reactive methods, that is, non-invasive methods for studying social processes available on the Internet.

The core competence which students will acquire is the knowledge of the various type of data available on the Internet, on the techniques for their collection, and of the capacity to actual use those techniques and certain tools for collecting data.

Programme (Brief Description of modules) and expected learning outcomes

Nowadays the Internet and the Web are more and more a place where people interact, discuss, communicate, and establish social and working relationships. In the last years a set of techniques have been developed for studying those social processes. Those techniques can be applied in many different type of studies and in different field of research like social sciences, social media research, internet psychology, marketing studies, and many others. Some of those techniques may be related to the observation of the behaviour of people, or on the collection of data on the Internet that may be recorded in different formats.

This course gives an introduction and an overview of non-reactive data collection methods, covering the following topics:

- Different capture techniques: cookies, log files, environment variables, time measurement, paradata, and so on.
- Introduction to unstructured, semi-structured, and structured data on the Web, and presentation of various data models and languages for data representation.
- Techniques and tools for collecting and querying data from the Web:
 - Tools for crawling the Web and transforming text data in structured data;
 - Techniques for collecting data from social networks using their web services interfaces;
 - Techniques for submitting queries on structured data sets.

Methodology

The course will be organized with a series of lectures.

- Resources:

Bibliography:

- Dietmar Janetzko. Non-reactive Data Collection on the Internet. The SAGE Handbook of Online Research Methods, Nigel Fielding & Raymond M. Lee & Grant Blank, 2008 <http://srmo.sagepub.com/view/the-sage-handbook-of-online-research-methods/n9.xml?rskey=G9FRT1>
- Bizer, Christian, Heath, Tom and Berners-Lee, Tim (2009) Linked Data - The Story So Far. Int. J. Semantic Web Inf. Syst., 5, (3), 1-22. <http://eolo.cps.unizar.es/docencia/doctorado/Articulos/LinkedData/bizer-heath-berners-lee-ijswis-linked-data.pdf>

Online resources:

- W3C Schools XML Tutorial <http://www.w3schools.com/xml/>
- W3C Schools RDF Tutorial <http://www.w3schools.com/rdf/>

Other on-line resources and links to papers containing interesting use-cases will be provided by the teacher during the course.

Evaluation System:

General Considerations:

Evaluation Criteria:

Students will be evaluated based on an exam which will evaluate their knowledge of the various techniques presented in the course.

Recommendation for second and following evaluations:

Employment Opportunities (optional)

The techniques and tools discussed in the course may be used in many field of applications, like for example on-line media research, digital marketing, and social science studies.