COURSE: <u>Social Media Analysis</u>

INSTRUCTORs: Roberto BASILI, Danilo CROCE
EMAIL: Dipartimento di Ingegneria dell'Impresa, Via del Politecnico 1. Tel 06 7259 7391,
    email : basili@info.uniroma2.it, croce@info.unirma2.it

WEB PAGE: http://directory.uniroma2.it/index.php/chart/dettagliDocente/3546,
http://directory.uniroma2.it/index.php/chart/dettagliDocente/8152

COURSE DESCRIPTION
The course provides an introduction to the adoption of Machine Learning methods in the analysis of Social Networks, both in terms of information access technologies as well as regarding the possible analytics functions over Social Networks, e.g. profiling, emergence of communities and recommending. In the course, we will overview advanced technologies for Information Retrieval over Social Networks. Then, their applications to more complex tasks such as Sentiment Analysis, Community Detection and Recommending will be discussed. Experimental activities will refer to standard benchmarking datasets as well as to specific tasks (e.g. Sentiment Analysis over Twitter in English and Italian).

LEARNING OUTCOMES
The course is aimed to provide students the following:
- Knowledge relevant paradigms and best practices in the area of Information Retrieval and Data Mining over Web data.
- Knowledge about the main problems in the processing of unstructured data in the Web, such as the contents of interactions and other activities within a Social Network, that are the focus of quantitative methods of statistical language processing and neural networks.
- Recognition of the different problems and opportunities inspired by the data arising in the streams generated by social networks, Web blogs and micro-blog platforms
- Knowledge about the technologies for modelling and solving decision-making and predictive processes over Social Web data.
- Ability to makes direct use of language processing and machine learning tools as applied to Social Networks analysis tasks, along the modelling (i.e. problem understanding and design of a ML solution) as well as experimental (i.e. performance measurement and optimization over known datasets) dimensions

METHODOLOGY
The methodological aspects of the course will be covered by a theoretical part (about 65-70% of the course covered by Roberto Basili), devoted to introducing relevant and widely adopted approaches to Information Retrieval, Semantic Search and Social Network Modeling. A more practical section of the course will concentrate on the application of the above models and methodologies to Sentiment Analysis, Community Detection and Recommending tasks over real datasets (about 30-35% of the course involving Danilo Croce).
The course will include some laboratory with specific problems proposed to students that, organized in groups and suitably monitored/assisted, will be asked to accomplish to the design,

development and measurement of focused Social Network analysis tasks (e.g. monitor the user appreciation of TV programs over twitter).

ASSESSMENT
Written exam; weighting: 70%
Project; weighting: 30%

OUTLINE
Section I: *Short Introduction to Information Management and Retrieval in the Web*.
Introduction to Social Network Analysis. Traditional Information Retrieval: document processing, indexing, ranking. Target tasks for Web IR. Ranking Models for the Web. Random walk models: Page Rank. Web Search Engines. Preference Learning in IR.

Section II: *Advanced Language Processing for Social Network Analysis*.
Statistical and Neural NLP. Tree kernels. Distributional methods for Natural Language Semantics. Latent Semantic kernels. Semantic Tree Kernels. Applications: Information Extraction; Question Answering; Open-domain Information Extraction.

Section III: *Social Media Analytics*.
Social Web. Graph-based algorithms for community detection. Introduction to Opinion Mining and Sentiment Analysis. Recommending Systems.

TEXTBOOKS
- **IR -** *Introduction to Information Retrieval*, Christopher D. Manning, Prabhakar Raghavan and Hinrich Schutze, Cambridge University Press. 2008. Find the book Home page HERE.
- **Social Media Analytics -** *Community Detection and Mining in Social Media*, Lei Tang, Huan Liu, Morgan & Claypool Publishers, 2010.

ADDITIONAL SUGGESTED READINGS
- **Social Media Analytics -** *Analyzing the Social Web*, Jennifer Golbeck, Elsevier, 2015.
- **ML e Web -** Bing Liu, *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*. 2nd Edition, July 2011, Springer.
- ***Machine Learning -** Pattern Recognition and Machine learning*, C. Bishop. Springer. 2006.
- **ML ed IR -** *Automatic Text Categorization: from Information Retrieval to Support Vector Learning*, Roberto Basili, Alessandro Moschitti, ARACNE Editore, 2005.

The course material will be made available during the course: slides, datasets, supplementary materials (Java code or Python scripts).
The dedicated Web site is at:
http://ai-nlp.info.uniroma2.it/basili/didattica/BigData2018/