

# CORSO DI *WEB MINING E RETRIEVAL* *- INTRODUZIONE AL CORSO -*

---

Corsi di Laurea in Informatica, Ing.  
Informatica, Ing. di Internet  
(a.a. 2015-2016)

Roberto Basili

# Overview

- WM&R: Motivazioni e prospettive
- Modalità di erogazione del Corso
- Prerequisiti
- Forma e struttura delle prove d'esame

# Obbiettivi del Corso

- Metodi per l'accesso ed il trattamento dell'informazione distribuita
- Componente Fondazionale:
  - Problemi di Ricerca e Classificazione dei Testi
  - Modelli Avanzati di *Machine Learning*
  - Advanced Language Processing
- Applicazioni:
  - *Enterprise Semantic Search, Opinion & Social Network Analysis, Cognitive Computing*
  - Annotazione Automatica di Testi, Musica ed Immagini

# Organizzazione del Corso:

## Sezione I: Advanced Machine Learning

Intro al ML. Supervised ed Unsupervised Learning

Algoritmi generativi: Hidden Markov Models

Teoria Statistica dell'Apprendimento e Kernel-based learning.

Support Vector Machine. Deep Neural Networks.

## Sezione II: Statistical Language Processing

Language Modeling and Sequence labeling

Statistical Parsing. Information Extraction. Question Answering

## • Sezione III: Web Mining & Retrieval

- Sistemi e Tecnologie per l'IR nel Web.

- Web Retrieval. Algoritmi di Ranking. Page Rank.

- Introduzione al Social Network Analysis, Opinion Mining.

# Organizzazione del Corso:

## Sezione I: Advanced Machine Learning

Intro al ML. Supervised ed Unsupervised Learning

Algoritmi generativi: Hidden Markov Models

Teoria Statistica dell'Apprendimento e Kernel-based learning.

Support Vector Machine. Deep Neural Networks.

## Sezione III: Web Mining & Retrieval

Sistemi e Tecnologie per l'IR nel Web.

Web Retrieval. Algoritmi di Ranking. Page Rank.

Introduzione al Social Network Analysis, Opinion Mining.

- Sezione II: Statistical Language Processing

- Language Modeling and Sequence labeling

- Statistical Parsing. Information Extraction. Question Answering

# Obbiettivi (2):

- Aspetti fondazionali
  - Modelli ed Algoritmi di Apprendimento Automatico
  - Ottimizzazione: Feature Selection, Dim. Reduction, Statistical Learning Theory
  - Modelli di IR, motori di ricerca
- Esercitazioni
  - Completamento della Teoria
  - Introduzioni a sistemi e tecnologie
    - Data Mining, Search Engines, Web crawler
  - Progetto Finale
    - Applicazioni di modelli avanzati di IR
    - ML for IR
    - Statistical Natural Language Processing

# Target

- Laurea Triennale in Informatica (o Ing. Inf. Aut., Int., Gest.)
- Prerequisiti:
  - Elementi di Analisi e Geometria
  - Elementi di Logica
  - Calcolo delle Probabilità e Statistica
  - Elementi di Basi di Dati e Sistemi Basati su Conoscenza
- Vedi Corsi della Laurea di Informatica:
  - ML (G. Gambosi)
  - IR (G. Amati)

# Orari

- **Lunedì' 14:00 - 16:00 AULA C2 -**  
Ed. Nuovo della Didattica, presso la Facolta' di Ingegneria
- **Mercoledì' 16:00 - 18:00 AULA C3 -**  
Ed. Nuovo della Didattica, presso la Facolta' di Ingegneria
- **Giovedì' 11:30 - 13:30 AULA C2 -**  
Ed. Nuovo della Didattica, presso la Facolta' di Ingegneria

**Ricevimento: LUNEDI' 13:15-14:00**

# Sito Web

- Didattica Web:
  - Ing.  
<http://didattica.uniroma2.it/informazioni/index/insegnamento/160975-Web-Mining-And-Retrieval>
  - Scienze:  
<http://didattica.uniroma2.it/informazioni/index/insegnamento/160974-Web-Mining-And-Retrieval>
- I materiali pubblicati sono sul sito:
  - [http://art.uniroma2.it/basili/didattica/WmIR\\_15\\_16/](http://art.uniroma2.it/basili/didattica/WmIR_15_16/)

# Sito del Corso

## Web Mining e Retrieval (a.a. 2015/16) Secondo Semestre

[Esci dai Frame](#)



Elenco dei File nel deposito



### Sommario Contenuti

1. [Novità](#)
2. [Programma del Corso](#)
3. [Testi di Riferimento](#)
4. [Link Utili](#)
5. [Diapositive delle lezioni](#) 
6. [Progetti ed Esercizi Proposti](#) 

# Testi Consigliati

- IR Introduction to Information Retrieval, Christopher D. Manning, Prabhakar Raghavan and Hinrich Schutze, Cambridge University Press. 2008. <http://nlp.stanford.edu/IR-book/>
- MLxIR. Automatic Text Categorization: From Information Retrieval to Support Vector Learning, R. Basili, A. Moschitti. ARACNE Ed., 2005
- Materiale complementare
  - Pattern Recognition and Data Mining, C. Bishop, 2006.
  - "Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data", Liu, Bing, Springer-Verlag, Series: Data-Centric Systems and Applications, Berlin, ISBN: 978-3-540-37881-5, 2007.
  - Articoli e Dispense
    - Dispense e Lucidi delle lezioni:  
[http://art.uniroma2.it/basili/didattica/WmIR\\_15\\_16/](http://art.uniroma2.it/basili/didattica/WmIR_15_16/)

# Organizzazione: Esami

- *Modalità del Corso*
  - *Full (9 crediti)*
    - *Informatica, Laurea Tecnologie di Internet*
  - *Light (6 crediti)*
    - *Laurea Ing. Informatica, Automatica e Gestionale*

# Esami Light

- Esonero MidTerm (intorno all'11 Aprile)
- Scritto finale (dopo il 10 Giugno 2013)
- Sezioni I, II e III
  
- E' facoltativo ( tirocinio?)
  - un progetto (ad esempio, una applicazione IR in ambiente mobile, Android)  
oppure
  - un approfondimento bibliografico (ricerca bibliografica su un arg. avanzato)

# Esami Full

- Esonero MidTerm (intorno all'11 Aprile)
- Scritto finale (dopo il 10 Giugno 2013)
- Sezioni I, II e III
  
- E' consigliato
  - un progetto (ad esempio, sperimentazione di modelli di ML, applicazioni IR in ambiente mobile Android, recommender system)  
oppure
  - un approfondimento bibliografico

# Esami (1)

- **Approfondimento Bibliografico**
  - assegnamento di un tema e della relativa bibliografia,
  - ricevimento dedicato (fuori degli orari delle lezioni),
  - produzione di una relazione finale (stile Tesi),
  - Esame: presentazione finale della tesina (poss. entro la fine del Corso)

# Esami (2)

- **Progetto** (max 2/3 persone)
  - Assegnamento di un progetto (stile tesi),
  - ricevimento dedicato (fuori degli orari delle lezioni),
  - produzione di un dimostrativo e di una breve relazione
  - Esame: Discussione del progetto + Demo

# Esempi di Progetti

- **Opinion Tracker.** Riconoscimento e tracking di opinion leaders in social networks
- **Multilingual Search Engine.** Multilingual extension of an enterprise search engine (Reveal)
- **Graph-based Optimizer.** Algoritmi graph-based per machine learning.
- **Playlist Generator.** Sviluppo di un sistema di classificazione e recommending su informazione musicale (audio files, Web pages, MIDI)
- **Community Detection** in Open Linked Data
- **Automatic Metadata creator.** Titolazione automatica di immagini e video

Domande?

# Action List

- Registrarsi al Corso presso Delphi e presso :
  - URL: <http://didattica.art.uniroma2.it/esami/>
- Definire i propri estremi e tipo di Corso (ad es. i CFU e o i Corsi già sostenuti) tramite il campo “Note”
- Verranno pubblicati:
  - Elenchi dei gruppi registrati
  - Progetti
  - Orari ricevimento per gli studenti che non seguono
  - Slide e materiali complementari (*in progress*)