

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

Corsi di Laurea in Informatica,
Ing. Gestionale, Ing. Informatica, Ing. di Internet
(a.a. 2021-2022)

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 e Deep Learning*
 - *Statistical Learning Theory*
 - *Kernel Machines*
 - *Artificial Neural Networks e Deep Learning*
 - Advanced Natural Language Processing
- Applicazioni:
 - *Textual Inference, Semantic Search, Question Answering*
 - *Opinion Mining & Sentiment Analysis*
 - Multimedia Data Mining: Machine Vision & Data Analytics

Organizzazione del Corso:

Sezione I: Advanced Machine Learning & Deep Learning

Intro ML. Supervised & Unsupervised Learning

Approcci probabilistici: Hidden Markov Models

Teoria Statistica dell'Apprendimento e Kernel-based learning.

Semi-supervised Methods.

Support Vector Machines.

Deep Learning with Neural Networks.

Sezione II: Natural Language Processing

Statistical Language Processing. Language Modeling.

Textual Inference Semantic Role Labeling.

• Sezione III: Web Mining and Advanced Analytics Applications

- Information Extraction. Question Answering.
- Opinion Mining & Sentiment Analysis.
- Visual Object Recognition and Automatic Image Captioning

Obbiettivi (2):

- Aspetti fondazionali
 - Modelli ed Algoritmi di Apprendimento Automatico
 - Ottimizzazione: *Feature Selection*, *Dimensionality Reduction*, *Statistical Learning Theory*
 - Modelli complessi di Data Analysis nel Web
- Esercitazioni
 - Completamento della Teoria
 - Introduzioni a sistemi e tecnologie
 - Data Mining sw, NLP sw, ML toolkit
 - Progetto Finale
 - Advanced Analytics applications
 - ML for IR and knowledge discovery
 - Statistical Natural Language Processing

Target

- Laurea Magistrale 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
- Corsi con argomenti correlati della Laurea Specialistica di Informatica:
 - *Machine Learning* (G. Gambosi)
 - *Natural Language Processing* (F.M. Zanzotto)
 - *Information Retrieval* (G. Gambosi, D. Croce)

Orari

- **Lunedì' 14:00 - 16:00 AULA 19**
presso la Macroarea di Scienze
- **Mercoledì' 9:30 - 11:30 AULA 13**
presso la Macroarea di Scienze
- **Giovedì' 11:30 - 13:30 AULA 13**
presso la Macroarea di Scienze

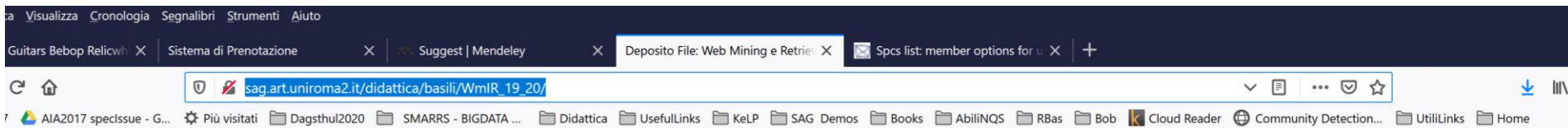
Ricevimento:

**ogni Giovedì dopo la lezione o
su richiesta anche *on-line* (MS Teams)**

Sito Web

- Didattica Web (Corso di Laurea in Informatica)
- I materiali pubblicati sono sul sito:
- http://sag.art.uniroma2.it/didattica/basili/WmlR_21_22/

Sito del Corso



Web Mining e Retrieval (a.a. 2019/20) Secondo Semestre

Esci dai Frame

Elenco dei File nel deposito



Sommario Contenuti

1. [Novita'](#)
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/>
- Deep Learning, Ian Goodfellow, Yoshua Bengio, Aaron Courville, MIT Press, 2016.
- 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://sag.uniroma2.it/basili/didattica/WmIR_21_22/

**WEB MINING & RETRIEVAL,
a.a. 2020-21**

Moduli	Argomenti	Lezioni
Introduzione agli elementi di base del ML	Introduzione all'algorithmica di ML. Probabilità e Metriche di similarità. Classificazione mediante algoritmi di base	Introduzione al Corso, ed al WM&R.
		Richiami ai metodi base di ML. Metodi Supervised vs. Unsupervised.
		Metodi probabilistici e generativi. Naive Bayes.
		Esercitazione. Decision Trees. Intro Weka (1)
Clustering and Probabilistic methods	Introduction to clustering algorithms. Generative Models. HMMs	Metodi Unsupervised. Clustering. Metodi Agglomerativi: K-means
		Esercitazione. Metodi avanzati di Clustering. Uso di Weka (2)
		Modelli di Linguaggio. Processi Markoviani.
		Modelli Generativi: HMM. (POS tagging)
From PAC learnability to SVM	PAC learnability. VC-dimension. SVMs. Kernel methods	PAC Learnability. Perceptron
		SVM. Hard Margin.
		Soft margin SVM. La Nozione di Kernel.
		Kernel polinomiali e RBF. Sequence & Tree Kernels. Use of Tree and Sequence Kernels in KELP
Advanced Text Processing and Statistical Natural Language Processing	Standard Natural Language Processing. Data and Content Analysis in NLP. Web Applications. Evaluation of Statistical NLP tools.	Richiami ai metodi di Elaborazione del Linguaggio Naturale: il TAL
		Risorse: Corpora, Lessici, Grammatiche e Basi di Conoscenza per il TAL.
		Content and semantics in Textual Data
		Esercitazione: SpaCy and NLP Evaluation.
Advanced ML topics: Neural Networks	Semi-supervised learning: ensemble methods, active learning, EM. <i>On-line learning: Passive-Aggressive.</i> Deep Neural network architectures. Application and Design Practices of NNs	Improving ML settings: Semisupervised Extensions
		<i>Improving ML settings: On-line Learning con esercitazione</i>
		From neural networks to deep learning: perceptrons and MLP
		Deep Learning over MLPs. Recurrent Neural Networks. Adopting Convolutional Neural Networks on images,
ML and Advanced IR topics	Dimensionality Reduction for IR. Distributional Semantic models for IR. Word Embeddings with NNs.	Latent Semantic Analysis. LSA. Query Operations/Expansion - Automatic Global Analysis - Thesaurus/SA
		Latent Semantic Kernels & Semantic Kernels. Wordspaces
		Word Embeddings through Neural Networks.
		Word Space Exercise
Social Media Analytics	IR in Social Media. Community detection. User profiling and Recommending. Sentiment and Emotion Analysis.	Introduzione al web and Social Medi Analysis
		Opinion Mining e Sentiment Analysis: the task
		Introduzione all'OM & SA: Twitter as a case study
		Development of a SA tool Social Media Analysis: Recommending

Organizzazione: Esami

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

Esami Light (6 CFU)

- Esonero MidTerm (intorno al 20 Aprile 2022)
- Scritto finale (intorno al 10 Giugno 2022)
- Sezioni I, II e III

- E' facoltativo (tirocinio?)
 - un progetto (ad esempio, una applicazione Web o mobile)
 - Deep Learning su immagini
 - Deep Learning for Fake News Detection
 - Sistemi di addestramento per interfacce vocali roboticheoppure
 - un approfondimento bibliografico (ricerca bibliografica su un arg. avanzato)

Esami Full (9 CFU)

- Obbligatoria la consegna di un certo numero di esercizi proposti a lezione
- Esonero MidTerm (intorno all'11 Aprile)
- Scritto finale (dopo il 10 Giugno 2021)
- Sezioni I, II e III

- E' consigliato
 - un progetto (ad esempio, sperimentazione di modelli di ML, applicazioni IR in ambiente mobile Android, sentiment analysis, deep learning, recommending)
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 opinioni in social networks sulla base di descrizioni tematiche
- **Multimedia Search Engine.** Enterprise search in picture and video repositories
- **Graph Neural Networks.** Algoritmi graph-based per machine learning e ragionamento automatico.
- **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 presso :
 - URL: <https://delphi.uniroma2.it/totem/jsp/>
- Definire i propri estremi e tipo di Corso (ad es. i CFU e o i Corsi già sostenuti) tramite il campo “Note”
- Verranno pubblicati:
 - Slide e materiali complementari (*in progress*)
 - Orari ricevimento per gli studenti che non seguono
 - Elenchi dei gruppi registrati
 - Progetti