

Principali informazioni sull'insegnamento	A.A. 2018-2019
Titolo insegnamento	Metodi per il Ritrovamento dell'Informazione
Corso di studio	Laurea Triennale in Informatica
Crediti formativi	9
Denominazione inglese	Information Retrieval
Obbligo di frequenza	No
Lingua di erogazione	Italiano

Docente responsabile	Nome Cognome	Indirizzo Mail
	Pasquale Lops	pasquale.lops@uniba.it
Luogo ed orario di ricevimento	Dipartimento di Informatica 7° Piano stanza 760	Martedì dalle 10 alle 12

Dettaglio credi formativi	Ambito disciplinare	SSD	Crediti
	INFORMATICO	ING-INF/05	9 (7T1+2T2)

Modalità di erogazione	
Periodo di erogazione	Primo semestre
Anno di corso	Terzo anno
Modalità di erogazione	Lezioni frontali Esercitazioni in Aula e Laboratorio

Organizzazione della didattica	
Ore totali	225
Ore di corso	86
Ore di studio individuale	139

Calendario	
Inizio attività didattiche	24 settembre 2018
Fine attività didattiche	11 gennaio 2019

Syllabus	
Prerequisiti	Matematica Discreta Linguaggi di Programmazione Calcolo delle Probabilità e Statistica Algoritmi e Strutture Dati
Risultati di apprendimento previsti (declinare rispetto ai Descrittori di Dublino) (si raccomanda che siano coerenti con i risultati di apprendimento del CdS, compreso i risultati di apprendimento trasversali)	<ul style="list-style-type: none"> • <i>Conoscenza e capacità di comprensione</i> Lo studente dovrà acquisire le competenze relative alla gestione di sorgenti di informazioni non strutturate. Lo strumento attraverso il quale raggiungere questo obiettivo è un approccio metodologico basato sulla conoscenza dei meccanismi che ispirano il progetto e l'implementazione dei sistemi di ritrovamento e filtraggio delle informazioni.

	<ul style="list-style-type: none"> • <i>Conoscenza e capacità di comprensione applicate</i> Lo studente dovrà acquisire le competenze necessarie per lo sviluppo e la realizzazione di sistemi di ricerca e filtraggio delle informazioni (motori di ricerca, recommender systems). • <i>Autonomia di giudizio</i> Lo studente deve dimostrare di aver acquisito una notevole autonomia di giudizio e di gestione delle problematiche relative all'uso delle metodologie di ricerca e filtraggio dell'informazione. • <i>Abilità comunicative</i> Lo studente sarà in grado di illustrare in modo appropriato le caratteristiche metodologiche e tecniche degli strumenti di ricerca e filtraggio dell'informazione. • <i>Capacità di apprendere</i> Lo studente dovrà mostrare di aver sviluppato capacità di apprendere e di orientarsi agilmente nelle problematiche che si presentano durante l'utilizzo delle tecnologie per la gestione di sorgenti di informazioni non strutturate.
<p>Contenuti di insegnamento</p>	<p>Accesso Intelligente all'Informazione e gestione dell'informazione non strutturata</p> <ul style="list-style-type: none"> • Text Mining: estrazione di conoscenza da informazioni non strutturate • Definizione di Text Mining e processo di Text Mining • Introduzione all'elaborazione del linguaggio naturale <p>Information Retrieval (IR)</p> <ul style="list-style-type: none"> • Architettura di un sistema di IR • Modelli di ritrovamento dell'informazione • Modello Booleano • Modello a Spazio Vettoriale • Integrazione di conoscenza lessicale: WordNet • Relevance Feedback e algoritmo di Rocchio • Link Analysis e Page Rank <p>Metriche per la valutazione dei sistemi di IR</p> <ul style="list-style-type: none"> • Cranfield Paradigm • Precision, Recall, F-measure • Valutazione del ranking: <ul style="list-style-type: none"> - Curve di precision-recall - Mean Average Precision (MAP) - normalized Discounted Cumulative Gain (nDCG) • Rank Correlation Metrics <ul style="list-style-type: none"> - Coefficienti di Spearman e Kendall Tau <p>Text Categorization (TC)</p> <ul style="list-style-type: none"> • Generalità sul problema della classificazione • Estensione del problema ai dati testuali: Text Categorization (TC)

	<ul style="list-style-type: none"> • Machine Learning for TC <ul style="list-style-type: none"> - Metodo di Rocchio - k-Nearest Neighbor (k-NN) - Naive Bayes - Metriche per la valutazione dei sistemi di TC <p>Accesso Personalizzato all'Informazione e Recommender Systems</p> <ul style="list-style-type: none"> • Information Retrieval vs Information Filtering (IF) • Architettura di un sistema di IF • Recommender Systems <ul style="list-style-type: none"> - Collaborative Filtering (CF) - Content-based Filtering • Metodologie e metriche per la valutazione di Recommender Systems <p>Semantic Web</p> <ul style="list-style-type: none"> • Introduzione • I Linguaggi del Semantic Web: RDF e SPARQL • Introduzione ai Linked Open Data <p>Esercitazioni: Apache Lucene/MyMediaLite</p>
--	---

Programma	
Testi di riferimento	<p>Ricardo Baeza-Yates and Berthier Ribeiro-Neto. Modern Information Retrieval: The Concepts and Technology Behind Search (Second edition). Addison Wesley. ISBN: 978-0-321-41691-9.</p> <p>Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütz. Introduction to Information Retrieval, Cambridge University Press, 2008. ISBN: 978-0521865719. Versione digitale: https://nlp.stanford.edu/IR-book/.</p> <p>Dietmar Jannach, Markus Zanker, Alexander Felfernig, Gerhard Friedrich. Recommender Systems An Introduction, Cambridge University Press, 2010. ISBN: 978-0521493369.</p>
Note ai testi di riferimento	<p>I libri di testo sono integrati con le slide e le dispense del docente e verranno forniti articoli scientifici selezionati dal docente ad integrazione del materiale didattico di riferimento. Tutto il materiale sarà reso disponibile sulla piattaforma di e-learning del Dipartimento di Informatica.</p>
Metodi didattici	<p>Le lezioni frontali (56 ore) saranno dedicate all'apprendimento dei modelli teorici e dei concetti di base coadiuvati da alcuni esempi.</p> <p>Le esercitazioni (30 ore) saranno dedicate all'esecuzione di esercizi in classe anche coinvolgendo direttamente gli studenti nella risoluzione degli stessi e alla realizzazione di semplici algoritmi per la ricerca e il filtraggio dell'informazione.</p>
Metodi di valutazione (indicare almeno la tipologia scritto, orale, altro)	<p>1) Appelli d'esame (al termine dell'insegnamento) - L'esame consta di una prova scritta che comprende sia domande teoriche che svolgimento di esercizi. La durata della prova varia in base alla tipologia di quesiti da risolvere, non</p>

	<p>superando comunque i 90 minuti. Durante la prova non è ammessa la consultazione di materiale didattico ed è ammesso l'eventuale uso di una calcolatrice.</p> <p>2) Prova intermedia (durante lo svolgimento dell'insegnamento) - E' prevista una prova scritta intermedia, articolata su una specifica parte del programma indicata dal docente prima dell'interruzione delle lezioni prevista da calendario accademico. Il superamento della prova intermedia esonera lo studente dallo svolgimento di una parte della prova d'esame, purché quest'ultima sia sostenuta nel primo appello di giugno.</p>
<p>Criteri di valutazione (per ogni risultato di apprendimento atteso su indicato, descrivere cosa ci si aspetta lo studente conosca o sia in grado di fare e a quale livello al fine di dimostrare che un risultato di apprendimento è stato raggiunto e a quale livello)</p>	<p>Nella prova scritta lo studente dovrà dimostrare di aver compreso il funzionamento degli algoritmi di ricerca e filtraggio dell'informazione. Inoltre, dovrà dimostrare di aver acquisito le competenze per la realizzazione di semplici strumenti per la gestione dell'informazione non strutturata.</p> <p>La valutazione della prova è espressa in trentesimi. La prova intermedia è superata con un minimo di 16/30, conseguito svolgendo correttamente almeno uno dei quesiti posti. La prova d'appello è superata con un minimo di 18/30. Per gli studenti non esonerati, il voto finale d'esame è quello riportato nella prova scritta d'appello. Per gli studenti esonerati il voto finale d'esame è la media delle valutazioni riportate nella prova intermedia ed in quella del primo appello di giugno.</p> <p>La determinazione del voto tiene conto dei seguenti aspetti:</p> <ol style="list-style-type: none"> 1) la correttezza della soluzione proposta; 2) la completezza della soluzione proposta; 3) la logica seguita dallo studente nel proporre la soluzione. <p>Per superare la prova d'esame o la prova intermedia, lo studente deve essere in grado di proporre una soluzione che soddisfi almeno l'aspetto 1). I voti superiori al minimo vengono attribuiti agli studenti la cui soluzione soddisfi anche gli altri aspetti.</p>
<p>Altro</p>	