



## Principali informazioni sull'insegnamento

Denominazione dell'insegnamento	<b>Metodi per il Ritrovamento dell'Informazione</b>	
Corso di studio	Informatica triennale (L31) – sede di Bari	
Anno Accademico	2023/24	
Crediti formativi universitari (CFU) / European Credit Transfer and Accumulation System (ECTS)	9 CFU	
Settore Scientifico Disciplinare	INF/01	
Lingua di erogazione	Italiano	
Anno di corso	Terzo	
Periodo di erogazione	1° semestre, le date esatte sono indicate nel manifesto/regolamento dell'anno accademico di riferimento	
Obbligo di frequenza	La frequenza è fortemente raccomandata	
Sito web del corso di studio	<a href="https://www.uniba.it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea/informatica-270/laurea-triennale-in-informatica-d.m.-270-1">https://www.uniba.it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea/informatica-270/laurea-triennale-in-informatica-d.m.-270-1</a>	

<b>Docente/i</b>	
Nome e cognome	Pasquale Lops
Indirizzo mail	<a href="mailto:pasquale.lops@uniba.it">pasquale.lops@uniba.it</a>
Telefono	080-5442276
Sede	Dipartimento di Informatica, Campus Universitario, Via E. Orabona 4, 70126, Bari. Ufficio n.760, 7° piano.
Sede virtuale	Piattaforma ADA <a href="https://elearning.uniba.it/">https://elearning.uniba.it/</a>
Sito web del docente	<a href="https://www.uniba.it/it/docenti/lops-pasquale">https://www.uniba.it/it/docenti/lops-pasquale</a>
Ricevimento	Martedì ore 10-12 oppure appuntamento da concordare per e-mail con il docente

<b>Syllabus</b>	
<b>Obiettivi formativi</b>	L'insegnamento si propone di introdurre: <ul style="list-style-type: none"><li>• i fondamentali teorici e pratici per la realizzazione di sistemi di ritrovamento e filtraggio delle informazioni.</li></ul>
<b>Prerequisiti</b>	Le seguenti conoscenze preliminari facilitano ed accelerano la comprensione degli argomenti dell'insegnamento: <ul style="list-style-type: none"><li>• da Programmazione: basi della programmazione imperativa;</li><li>• da Linguaggi di Programmazione: analisi lessicale e sintattica;</li><li>• da Matematica Discreta: teoria degli insiemi, relazioni, spazi vettoriali, operazioni su vettori e matrici.</li><li>• Da Calcolo delle Probabilità e Statistica: probabilità, probabilità condizionate, teorema di Bayes</li></ul>
<b>Contenuti di insegnamento (Programma)</b>	La numerazione degli argomenti ne indica l'ordine di presentazione e di studio. <ol style="list-style-type: none"><li>1. Accesso Intelligente all'Informazione e gestione dell'informazione non strutturata<ul style="list-style-type: none"><li>• Text Mining: estrazione di conoscenza da informazioni non strutturate</li><li>• Definizione di Text Mining e processo di Text Mining</li><li>• Introduzione all'elaborazione del linguaggio naturale</li></ul>Lezioni frontali: 3 ore</li><li>2. Information Retrieval (IR)<ul style="list-style-type: none"><li>• Architettura di un sistema di IR</li><li>• Modelli di ritrovamento dell'informazione</li><li>• Modello Booleano</li><li>• Modello a Spazio Vettoriale</li></ul></li></ol>



	<ul style="list-style-type: none"><li>• Integrazione di conoscenza lessicale: WordNet</li><li>• Relevance Feedback e algoritmo di Rocchio</li><li>• Link Analysis e Page Rank</li><li>• Apache Lucene</li></ul> <p>Lezioni frontali: 18 ore Esercitazioni: 13 ore</p> <p>3. Metriche per la valutazione dei sistemi di IR</p> <ul style="list-style-type: none"><li>• Cranfield Paradigm</li><li>• Precision, Recall, F-measure</li><li>• Valutazione del ranking:<ul style="list-style-type: none"><li>○ Curve di precision-recall</li><li>○ Mean Average Precision (MAP)</li><li>○ normalized Discounted Cumulative Gain (nDCG)</li></ul></li><li>• Rank Correlation Metrics<ul style="list-style-type: none"><li>○ Coefficienti di Spearman e Kendall Tau</li></ul></li></ul> <p>Lezioni frontali: 13 ore Esercitazioni: 5 ore</p> <p>4. Text Categorization (TC)</p> <ul style="list-style-type: none"><li>• Generalità sul problema della classificazione</li><li>• Estensione del problema ai dati testuali: Text Categorization (TC)</li><li>• Machine Learning for TC<ul style="list-style-type: none"><li>○ Metodo di Rocchio</li><li>○ k-Nearest Neighbor (k-NN)</li><li>○ Naive Bayes</li><li>○ Metriche per la valutazione dei sistemi di TC</li></ul></li></ul> <p>Lezioni frontali: 8 ore Esercitazioni: 6 ore</p> <p>5. Accesso Personalizzato all'Informazione e Recommender Systems</p> <ul style="list-style-type: none"><li>• Information Retrieval vs Information Filtering (IF)</li><li>• Architettura di un sistema di IF</li><li>• Recommender Systems<ul style="list-style-type: none"><li>○ Collaborative Filtering (CF)</li><li>○ Content-based Filtering</li></ul></li><li>• Metodologie e metriche per la valutazione di Recommender Systems</li></ul> <p>Lezioni frontali: 14 ore Esercitazioni: 6 ore</p>
<b>Testi di riferimento</b>	<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: <a href="https://nlp.stanford.edu/IR-book/">https://nlp.stanford.edu/IR-book/</a></p> <p>Dietmar Jannach, Markus Zanker, Alexander Felfernig, Gerhard Friedrich. Recommender Systems An Introduction, Cambridge University Press, 2010. ISBN: 978-0521493369.</p> <p>Nella sezione seguente è specificata in dettaglio la corrispondenza tra gli argomenti del programma ed i capitoli dei testi di riferimento. Gli studenti che lo desiderano possono ottenere i testi in prestito dalla Biblioteca. Si suggerisce di verificare l'eventuale disponibilità dei testi mediante il Sistema</p>



	Bibliotecario di Ateneo <a href="https://opac.uniba.it/easyweb/w8018/index.php?">https://opac.uniba.it/easyweb/w8018/index.php?</a> e contattare la biblioteca per concordare il prestito.		
<b>Note ai testi di riferimento</b>	<p>Si specificano di seguito, per ogni argomento del programma, i capitoli dei testi dai quali studiare.</p> <ol style="list-style-type: none"> <li>1. Accesso Intelligente all'Informazione e gestione dell'informazione non strutturata [Baeza-Yates] Capitolo 1, Capitolo 2 (sezione 2.1, 2.2)</li> <li>2. Information Retrieval [Baeza-Yates] Capitolo 3 (sezione 3.1, 3.2), Capitolo 5 (sezione 5.1, 5.2, 5.3) [Manning] Capitolo 1, 2, 6, 9, 21</li> <li>3. Metriche per la valutazione dei sistemi di IR [Baeza-Yates] Capitolo 4 (sezione 4.1, 4.2, 4.3, 4.4) [Manning] Capitolo 8</li> <li>4. Text Categorization [Baeza-Yates] Capitolo 8 (sezione 8.1, 8.2, 8.4, 8.5, 8.6) [Manning] Capitolo 13, 14</li> <li>5. Accesso Personalizzato all'Informazione e Recommender Systems [Jannach] Capitolo 1, 2, 3, 7</li> </ol> <p>Per ogni argomento è inoltre disponibile una dispensa, corrispondente ai contenuti proiettati in aula durante le lezioni, che non è sostitutiva dei testi di riferimento. Elenco del materiale disponibile sulla piattaforma di e-learning ADA <a href="https://elearning.di.uniba.it">https://elearning.di.uniba.it</a>:</p> <ul style="list-style-type: none"> <li>• dispense in formato pdf degli argomenti del programma;</li> <li>• esercizi con relativo svolgimento;</li> <li>• tracce di esempio delle prove d'esame, alcune delle quali con relative soluzioni</li> <li>• tracce di esempio della prova in itinere.</li> </ul>		
<b>Organizzazione della didattica</b>			
<b>Ore</b>			
Totali	Didattica frontale	Laboratorio ed esercitazioni	Studio individuale
225 ore	56 ore	30 ore	139 ore
<b>CFU/ETCS</b>			
9 CFU	7 CFU	2 CFU	
<b>Metodi didattici</b>	<ul style="list-style-type: none"> <li>• Lezioni frontali condotte con l'ausilio di trasparenze rese disponibili tramite la piattaforma di e-learning prima delle lezioni</li> <li>• Esercitazioni guidate in aula dedicate all'esecuzione di esercizi in classe anche coinvolgendo direttamente gli studenti</li> <li>• Esercitazioni guidate per la realizzazione di semplici sistemi per la ricerca e il filtraggio dell'informazione.</li> </ul>		
<b>Risultati di apprendimento previsti</b>			
<b>Conoscenza e capacità di comprensione</b>	<ul style="list-style-type: none"> <li>• 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.</li> </ul>		



<b>Conoscenza e capacità di comprensione applicate</b>	<ul style="list-style-type: none"><li>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).</li></ul>
<b>Competenze trasversali</b>	<p><b>Autonomia di giudizio</b></p> <ul style="list-style-type: none"><li>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.</li></ul> <p><b>Abilità comunicative</b></p> <ul style="list-style-type: none"><li>Lo studente sarà in grado di illustrare in modo appropriato le caratteristiche metodologiche e tecniche degli strumenti di ricerca e filtraggio dell'informazione.</li></ul> <p><b>Capacità di apprendere in modo autonomo</b></p> <ul style="list-style-type: none"><li>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.</li></ul>

<b>Valutazione</b>	
<b>Modalità di verifica dell'apprendimento</b>	<p><b><u>Prova d'esame</u></b></p> <ul style="list-style-type: none"><li>L'esame consta di una prova scritta che comprende sia domande teoriche che svolgimento di esercizi.</li></ul> <p>Regolamentazione della prova d'esame:</p> <ul style="list-style-type: none"><li>durata della prova: varia tra 60 e 90 minuti, in base alla tipologia di quesiti da risolvere;</li><li>materiale consultabile: nessuno; ammesso l'uso di una calcolatrice;</li><li>la valutazione è espressa in trentesimi;</li><li>comunicazione dei risultati: piattaforma esse3.</li></ul> <p><b><u>Prova intermedia</u></b></p> <p>È una prova scritta, <b>NON</b> obbligatoria, riguardante una specifica parte del programma indicata dal docente in prossimità dell'interruzione delle lezioni stabilita da Regolamento Didattico e Manifesto degli Studi.</p> <p>Regolamentazione della prova intermedia:</p> <ul style="list-style-type: none"><li>riservata solo agli studenti in corso</li><li>durata della prova: come prova d'esame;</li><li>materiale consultabile: come prova d'esame;</li><li>il superamento della prova intermedia esonera lo studente dallo svolgimento di una parte della prova d'esame (relativa a quanto già svolto nella prova intermedia), purché l'esame sia sostenuto nel <b>primo</b> appello;</li><li>la valutazione finale sarà ottenuta dalla media delle valutazioni conseguite nella prova intermedia ed in quella del primo appello;</li><li>l'obiettivo della prova intermedia è quello di dare un riscontro allo studente sul proprio livello di comprensione degli argomenti dell'insegnamento.</li></ul> <p><b><u>Simulazione prova d'esame</u></b></p> <p>Al termine delle lezioni si terrà una simulazione della prova d'esame, che consisterà nell'esecuzione di esercizi tratti da prove d'esame precedenti. Gli esercizi saranno svolti in autonomia dagli studenti nel tempo solitamente concesso in sede d'esame in modo da avere un riscontro della propria preparazione <i>prima</i> dell'esame. Al termine della prova il docente mostrerà le relative soluzioni.</p>
Criteria di valutazione	<ul style="list-style-type: none"><li><i>Conoscenza e capacità di comprensione:</i><ul style="list-style-type: none"><li>Capacità di descrivere con rigore i modelli di ritrovamento delle informazioni</li></ul></li></ul>



	<ul style="list-style-type: none"><li>○ Capacità di definire i limiti dei diversi modelli di ritrovamento, i punti di forza e di debolezza</li><li>● <i>Conoscenza e capacità di comprensione applicate:</i><ul style="list-style-type: none"><li>○ Capacità di proporre il miglior modello di ritrovamento in base al problema e al dominio specifici</li><li>○ Capacità di implementare praticamente e testare il modello di ritrovamento per comprenderne la qualità in base a diversi parametri di qualità</li></ul></li><li>● <i>Autonomia di giudizio:</i><ul style="list-style-type: none"><li>○ Modalità adottate per la verifica delle soluzioni proposte.</li></ul></li><li>● <i>Abilità comunicative:</i><ul style="list-style-type: none"><li>○ Chiarezza nella descrizione delle soluzioni proposte ai quesiti.</li></ul></li><li>● <i>Capacità di apprendere:</i><ul style="list-style-type: none"><li>○ Capacità di tradurre i requisiti di alto livello in documento di specifica e capacità di produrre una descrizione del progetto a diverse tipologie di audience.</li></ul></li></ul>
Criteri di misurazione dell'apprendimento e di attribuzione del voto finale	<p>Ad ogni quesito da svolgere sarà attribuito un punteggio, assicurando che la somma dei punteggi sia pari a 30.</p> <p>La determinazione del voto attribuito alla risposta data ad ogni quesito tiene conto dei seguenti criteri:</p> <ol style="list-style-type: none"><li>1) la correttezza della soluzione proposta;</li><li>2) la completezza della soluzione proposta;</li><li>3) la logica seguita dallo studente nel proporre la soluzione.</li></ol> <p>Per ottenere una valutazione sufficiente, la risposta fornita al quesito deve soddisfare almeno il criterio 1). I voti superiori alla sufficienza sono attribuiti se le risposte fornite soddisfano anche gli altri criteri.</p> <p>Il voto finale è dato dalla somma dei punteggi ottenuti sui singoli quesiti.</p> <p><b><u>Prova intermedia</u></b></p> <p>La valutazione della prova è espressa in trentesimi.</p> <p>La prova intermedia si intende superata quando il voto conseguito è maggiore o uguale a 16.</p> <p>L'esito della prova intermedia concorre alla valutazione finale come descritto nel riquadro "Modalità di verifica dell'apprendimento".</p> <p><b><u>Prova d'esame</u></b></p> <p>La valutazione della prova è espressa in trentesimi.</p> <p>La prova d'esame si intende superata quando il voto conseguito è maggiore o uguale a 18.</p> <p>Attribuzione della lode: la lode è attribuita quando la logica seguita dallo studente nel proporre la soluzione o la dimostrazione formale evidenzia particolari capacità di astrazione, ragionamento per analogia, creatività.</p>
Altro	<p>Si suggerisce agli studenti di affidarsi esclusivamente alle informazioni/comunicazioni fornite sui siti ufficiali del Dipartimento di Informatica, ovvero sui gruppi social solo se costituiti e amministrati esclusivamente dai docenti dei relativi insegnamenti:</p> <ul style="list-style-type: none"><li>● <a href="https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea/corsi-di-laurea">https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea/corsi-di-laurea</a></li><li>● <a href="https://www.uniba.it/it/ricerca/dipartimenti/informatica">https://www.uniba.it/it/ricerca/dipartimenti/informatica</a></li><li>● <a href="https://elearning.uniba.it/">https://elearning.uniba.it/</a></li></ul> <p>I programmi degli insegnamenti sono disponibili al seguente URL:</p> <ul style="list-style-type: none"><li>● <a href="https://programmi.di.uniba.it/">https://programmi.di.uniba.it/</a></li></ul> <p>Le informazioni che tutti gli studenti dovrebbero conoscere sono scritte nei Regolamenti Didattici e Manifesti degli studi disponibili sul sito:</p> <ul style="list-style-type: none"><li>● <a href="https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea/corsi-di-laurea">https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea/corsi-di-laurea</a></li></ul>



*Si suggerisce agli studenti di diffidare dalle informazioni e dei materiali circolanti su siti o gruppi social non ufficiali, poiché spesso sono risultati non affidabili, non corretti o incompleti. Per ogni dubbio, fissare un incontro con il docente secondo le modalità previste per il ricevimento.*

Link all'insegnamento sulla piattaforma e-learning del dipartimento ADA:

<https://elearning.uniba.it/>

**Suggerimenti per una corretta preparazione, in particolare per studenti non frequentanti:**

- 1) Studiare gli argomenti teorici nell'ordine suggerito prima di cimentarsi negli esercizi. L'ordine con il quale sono presentati gli argomenti è fondamentale per la comprensione degli stessi. Tentare di risolvere gli esercizi apprendendo solo dalle tracce di esempio e relative soluzioni non consente di acquisire le basi culturali per poter superare l'esame.  
Comprendere perché si arriva ad una certa soluzione è molto più importante che svolgere molti esercizi senza comprenderne del tutto le basi teoriche.
- 2) Svolgere gli esercizi, dei quali sono fornite le soluzioni, disponibili su ADA, in modo da confrontare le proprie soluzioni con quelle fornite dal docente.
- 3) Svolgere gli esercizi tratti da prove d'esame per valutare la propria preparazione.
- 4) Contattare il docente per chiedere eventuali chiarimenti e/o verificare le proprie soluzioni agli esercizi svolti in autonomia prima di iscriversi alla prova d'esame.
- 5) In caso di mancato superamento della prova, contattare il docente per la revisione del proprio elaborato e la comprensione degli errori commessi.