

## Syllabus Attività Formativa

<b>Anno Offerta</b>	2023
<b>Corso di Studio</b>	B33 - DATA SCIENCE
<b>Regolamento Didattico</b>	B33-22-22
<b>Percorso di Studio</b>	GEN - GENERICO
<b>Insegnamento/Modulo</b>	A423714 - WEB INTELLIGENCE -
<b>Attività Formativa Integrata</b>	-
<b>Partizione Studenti</b>	-
<b>Periodo Didattico</b>	S1 - Primo Semestre
<b>Sede</b>	
<b>Anno Corso</b>	2
<b>Settore</b>	ING-INF/05 - SISTEMI DI ELABORAZIONE DELLE INFORMAZIONI
<b>Tipo attività Formativa</b>	C - Affine/Integrativa
<b>Ambito</b>	70318 - Attività formative affini o integrative
<b>CFU</b>	6.0
<b>Ore Attività Frontali</b>	48.0
<b>AF_ID</b>	178681

<b>Tipo Testo</b>	<b>Codice Tipo Testo</b>	<b>Num. Max. Caratteri</b>	<b>Ob bl.</b>	<b>Testo in Italiano</b>	<b>Testo in Inglese</b>

<b>Lingua insegnamento</b>	LINGUA_INS	100	Sì	Inglese	English
<b>Contenuti</b>	CONTENUTI	2000	Sì	<p>Il corso fornisce un'introduzione ai metodi per indagare il comportamento degli utenti su siti web e sui social network. In particolare, verranno affrontate le tecniche per la valutazione dei contenuti, la quantificazione della somiglianza ed il posizionamento dei risultati di ricerca, la gestione della conoscenza in applicazioni web-based, il web semantico e le ontologie, dei dati accessibili per la condivisione e la collaborazione, le tecniche linguistiche e statistiche per il text mining e l'analisi dei contenuti ed i sistemi di raccomandazione semantica.</p> <p>In particolare saranno approfonditi i seguenti temi:</p> <ul style="list-style-type: none"> <li>-Il recupero di informazioni e la ricerca sul web;</li> <li>-La scansione del web e l'automazione del processo di aggiornamento;</li> <li>-L'estrazione dei dati strutturati da sorgenti eterogenee;</li> <li>-L'analisi del sentimento, la formazione e l'estrazione opinioni;</li> <li>-L'estrazione di informazioni sull'utilizzo del web;</li> </ul>	<p>The course provides an introduction to methods for investigating user's behavior on web sites and social networks. In particular, techniques for content recommendation, similarity quantification and search result ranking will be addressed, knowledge management in web-based applications, semantic web and ontologies, open linked data for sharing and collaboration, linguistic and statistical techniques for text mining and content analysis, semantic recommender systems. In particular, the following topics will be studied in depth :</p> <ul style="list-style-type: none"> <li>-Information retrieval and web search;</li> <li>-Web Crawling and automation;</li> <li>-Structured Data Extraction;</li> <li>-Sentiment Analysis and Opinion Mining;</li> <li>-Web Usage Mining;</li> <li>-Recommendation Systems;</li> <li>-Mining Social-Network Graphs.</li> </ul>

				-I sistemi di gradimento (recommendation); -Estrazione dei grafici dai social network.	
<b>Testi di riferimento</b>	TESTI_RIF		Sì	<p>Jure Leskovec, Anand Rajaraman, Jeff Ullman, Mining Massive Datasets (2nd Edition), Cambridge University Press, 2015 ISBN 978-1107077232</p> <p>Bing Liu , Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data Springer 2011, ISBN 978- 3642194597</p> <p>Charu C. Aggarwal, Data Mining – The Textbook, Springer ISBN 2015 978-3-319-14141-1</p> <p>Mark Newman, Networks, Oxford University Press, 2018, ISBN 9780192527493</p> <p>Andrea Isoni, Machine Learning for the Web, PACKT Publishing, 2016, ISBN 978-1-78588- 660-7</p> <p>Albert-László Barabási, Network Science, Cambridge University Press, 2016 , ISBN 9781107076266</p>	<p>Jure Leskovec, Anand Rajaraman, Jeff Ullman, Mining Massive Datasets (2nd Edition), Cambridge University Press, 2015 ISBN 978-1107077232</p> <p>Bing Liu , Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data Springer 2011, ISBN 978- 3642194597</p> <p>Charu C. Aggarwal, Data Mining – The Textbook, Springer ISBN 2015 978-3-319-14141-1</p> <p>Mark Newman, Networks, Oxford University Press, 2018, ISBN 9780192527493</p> <p>Andrea Isoni, Machine Learning for the Web, PACKT Publishing, 2016, ISBN 978-1-78588- 660-7</p> <p>Albert-László Barabási, Network Science, Cambridge University Press, 2016 , ISBN 9781107076266</p>
<b>Obiettivi formativi</b>	OBIETT_FORM	3000	Sì	<p>Gli obiettivi principali del corso sono:</p> <ul style="list-style-type: none"> <li>• Spiegare come funzionano i motori di</li> </ul>	<p>The main objectives of the course are to:</p> <ul style="list-style-type: none"> <li>• Explain the connection of information</li> </ul>

			<p>ricerca e come avviene il recupero delle informazioni e come funzionano</p> <ul style="list-style-type: none"> <li>• Fornire esempi di come l'analisi del sentimento può essere applicata a ai social e ad applicazioni aziendali, sapere come identificare l'accadimento di eventi attraverso l'analisi dei social.</li> <li>• Presentare diverse applicazioni per l'analisi del sentimento e della formazione delle opinioni</li> <li>• Definire i principali sistemi di raccomandazione e spiegare il funzionamento esempi di applicazioni reali</li> <li>• Dimostrare come l'estrazione di grafici può essere applicato ai social network e fornire esempi di problemi che possono essere risolti con queste tecniche.</li> <li>• Presentare metodi avanzati per la scansione del web e l'estrazione di dati strutturati.</li> <li>• Sapere analizzare i registry delle attività(logs) generati dagli utenti nei registri del server Web può portare ad una migliore comprensione dei siti Web e dell'utente.</li> <li>• Individuare le criticità nell'estrazione della conoscenza dai dati.</li> </ul> <p>Gli studenti impareranno come la semantica e i data accessibili vengono utilizzati per estrarre, rappresentare e fornire conoscenze</p>	<p>retrieval with search engines and how they work</p> <ul style="list-style-type: none"> <li>• Provide examples of how sentiment analysis can be applied on multiple social and business applications, know how to identify events through the analysis of social networks</li> <li>• Present the different applications of sentiment analysis and opinion mining</li> <li>• Define the main types of recommendation systems, explain how recommendation systems work and give examples of real world applications</li> <li>• Demonstrate how graph mining can be applied on social networks and provide examples of problems that can be solved with these techniques.</li> <li>• Present advanced methods for web crawling and structured data extraction.</li> <li>• Know how the analysis of user-generated data in web server logs can lead to improving web sites and understanding the users.</li> <li>• Explain the challenges of extracting knowledge from data.</li> </ul> <p>Students will learn how semantics and open data are used to extract, represent and provide knowledge on the web, as well as</p>
--	--	--	---	--

				<p>sul web, nonché come possono essere utilizzati in applicazioni intelligenti; formeranno la capacità di utilizzare le tecnologie semantiche e i dati collegati aperti per analizzare i contenuti web non strutturati e costruire sistemi di raccomandazione semantica e altre soluzioni intelligenti e capaci di definire ontologie che possono essere estese e riutilizzate da applicazioni di altre persone.</p>	<p>how they can be used in intelligent applications. Ability to use semantic technologies and open linked data to analyze unstructured web content and build semantic recommender systems and other intelligent solutions. Ability to define ontologies that may be extended and reused by other peoples' applications.</p>
<b>Prerequisiti</b>	PREREQ	2000	Sì	<p>Programmazione Statistica</p>	<p>Programming Statistics</p>
<b>Metodi didattici</b>	METODI_DID	2000	Sì	<p>Lezioni teoriche ed esercitazioni pratiche in laboratorio, utilizzando Python e applicazioni ad hoc. Le lezioni di laboratorio possono essere seguite individualmente utilizzando il materiale disponibile al Corso.</p>	<p>Theoretical lessons and practical exercises in the laboratory, using Python and ad hoc applications. Lab classes can be taken individually using the material which is available at course.</p>
<b>Altre informazioni</b>	ALTRO	2000	Sì	<p>Nessuna</p>	<p>None</p>
<b>Modalità di verifica dell'apprendimento</b>	MOD_VER_AP PR	3000	Sì	<p>Il raggiungimento degli obiettivi del corso è accertato attraverso una prova scritta (un progetto) e una discussione del progetto. L'esame orale consiste in domande aperte che consentono di verificare le conoscenze teoriche. I progetti e le loro discussioni</p>	<p>The achievement of the course objectives is assessed through a written examination (a project) and a discussion of the project. The oral exam consists of open questions that allow to test the theoretical knowledge. The projects and their discussions will be used to</p>

			<p>saranno utilizzati per valutare la capacità degli studenti di utilizzare i metodi presentati nel corso e di implementare gli strumenti basati su di essi.</p> <p>Gli argomenti per i progetti verranno assegnati su richiesta e saranno in parte basati su esercitazioni pratiche proposte durante le lezioni di laboratorio (per gli studenti frequentanti il corso). I progetti possono essere completati in piccoli gruppi, a condizione che i singoli contributi possano essere chiaramente identificati.</p> <p>I progetti contribuiranno al voto finale e saranno valutati in base a correttezza, efficienza, documentazione del codice e qualità del rapporto e capacità degli studenti di discutere l'implementazione e la teoria pertinente. Nel caso di progetti di gruppo, il voto è individuale e tiene conto anche del contributo, dichiarato o valutato al momento della discussione, e della dimensione del gruppo.</p> <p>Il progetto dovrà essere consegnato prima della prova orale e sarà discusso nell'ambito della prova orale.</p> <p>L'elaborato contribuisce alla valutazione del 50% del voto dell'esame finale.</p>	<p>assess the ability of the students in using the methods presented in the course and implementing tools based on them.</p> <p>The topics for the projects will be assigned on request and will be partially based on practical exercises proposed during lab classes (for students attending the course). Projects can be completed in small groups, provided that individual contributions can be clearly identified.</p> <p>Projects will contribute to the final grade and will be graded based on correctness, efficiency, code documentation and report quality, and capability of the students to discuss the implementation and the relevant theory. In the case of group projects, the grade is individual and also accounts for contribution, either declared or assessed at discussion time, and group size.</p> <p>The project must be delivered before the oral exam and will be discussed as part of the oral exam. The project contributes to the overall evaluation of the exam with a weight of 50%.</p>
--	--	--	--	--

<b>Programma esteso</b>	PROGR_EST		Sì <ol style="list-style-type: none"> <li>1. Recupero delle informazioni e ricerca sul web             <ol style="list-style-type: none"> <li>a. concetti di base del recupero delle informazioni</li> <li>b. misure di pertinenza</li> <li>c. misure di valutazione</li> <li>d. elaborazione preventive di testi e pagine Web</li> <li>e. indice invertito</li> <li>f. Indicizzazione semantica latente</li> </ol> </li> <li>2. Scansione del Web ed automazione</li> <li>3. Estrazione di dati strutturati</li> <li>4. Analisi del sentimento ed analisi delle opinioni             <ol style="list-style-type: none"> <li>a. applicazioni</li> <li>b. il problema della determinazione del Sentimento</li> <li>c. classificazione del sentimento espresso nel document</li> <li>d. soggettività dell'espressione scritta e classificazione dei sentimenti</li> </ol> </li> <li>5. Estrazione dei dati di utilizzo del Web             <ol style="list-style-type: none"> <li>a. raccolta e pretrattamento dei dati</li> <li>b. modellazione dei dati</li> </ol> </li> <li>6. Sistemi di valutazione             <ol style="list-style-type: none"> <li>a. concetti di base (matrice di utilità, Long Tail, applicazioni)</li> <li>b. valutazioni basate sul contenuto</li> <li>c. filtro collaborativo</li> </ol> </li> </ol>	<ol style="list-style-type: none"> <li>1. Information retrieval and web search             <ol style="list-style-type: none"> <li>a. Basic concepts of information retrieval</li> <li>b. Relevance Feedback</li> <li>c. Evaluation Measures</li> <li>d. Text and Web Page Pre Processing</li> <li>e. Inverted Index</li> <li>f. Latent Semantic Indexing</li> </ol> </li> <li>2. Web Crawling and automation</li> <li>3. Structured Data Extraction</li> <li>4. Sentiment Analysis and Opinion Mining             <ol style="list-style-type: none"> <li>a. Applications</li> <li>b. The problem of Sentiment Analysis</li> <li>c. Document Sentiment Classification</li> <li>d. Sentence Subjectivity and Sentiment Classification</li> </ol> </li> <li>5. Web Usage Mining             <ol style="list-style-type: none"> <li>a. Data Collection and Pre-Processing</li> <li>b. Data Modeling</li> </ol> </li> <li>6. Recommendation Systems             <ol style="list-style-type: none"> <li>a. Basic Concepts (Utility Matrix, Long Tail, Applications)</li> <li>b. Content-Based Recommendations</li> <li>c. Collaborative Filtering</li> <li>d. Dimensionality Reduction</li> </ol> </li> <li>7. Mining Social-Network Graphs             <ol style="list-style-type: none"> <li>a. Social Networks as Graphs</li> <li>b. Clustering of Social Network Graphs</li> </ol> </li> </ol>
-------------------------	-----------	--	--	--

			<ul style="list-style-type: none"><li>d. riduzione delle dimensioni</li><li>7. creazione dei grafici nelle reti social</li><li>a. i social network come grafici</li><li>b. aggregazione di grafici</li><li>c. rilevazione di comunità</li><li>d. rilevazione di eventi nei social network</li></ul>	<ul style="list-style-type: none"><li>c. Community Detection</li><li>d. Event Detection in Social Networks</li></ul>
--	--	--	---	--