



Main information about the course

Title of the course	Natural Language Processing	
Degree Course	Master's Degree in Computer Science (LM-18)	
Academic year	2022/23	
University credits (CFU) / European Credit Transfer and Accumulation System (ECTS)	6 credits	
Scientific Disciplinary Sector	ING-INF/05	
Language	English	
Course year	First	
Period	2nd semester, the exact dates are indicated in the manifesto/regulation of the academic year of reference	
Attendance	Attendance is strongly recommended	
Course website	https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea/computer-science/computer-science	

Course Staff

Name and surname	Marco de Gemmis
E-mail address	marco.degemmis@uniba.it
Telephone	Tel. 080-5443283
Address	Department of Computer Science, Via Orabona 4, 70125, Bari. Room no. 761, 7th floor.
Lecturer's website	https://www.uniba.it/it/docenti/degemmis-marco
Name and surname	Pierpaolo Basile
E-mail address	pierpaolo.basile@uniba.it
Telephone	080-5442301
Address	Department of Computer Science, Via Orabona 4, 70125, Bari. Room n.758, 7th floor.
Lecturer's website	http://www.di.uniba.it/~swap/index.php?n=Membri.Basile
Student Reception	Appointment to be agreed by email with the lecturers
Virtual location	ADA Platform - https://elearning.di.uniba.it

Syllabus

Learning objectives	<ul style="list-style-type: none"> • Knowledge of the theoretical foundations of Natural Language Processing (NLP); • Knowledge of practical issues for the development of NLP systems.
Prerequisites	<p>The following preliminary knowledge facilitates the understanding of the topics of the course:</p> <ul style="list-style-type: none"> • Basics of: Probability theory, programming
Contents (Program)	<p>The numbering of the topics indicates the order of presentation and study. In addition, the index of each topic is useful to find related content in the reference texts.</p> <ol style="list-style-type: none"> 1. Introduction to NLP and computational linguistics (lecture: 4 hours). 2. Lexical level: tokenization, regular expressions, normalization of the text (lecture: 4 hours). 3. Language models: n-grams, probability of a sentence, evaluation of language models (lecture: 6 hours). 4. Classification of documents and applications (lecture: 5 hours). 5. Part-of-Speech (PoS) Tagging: approaches for PoS tagging, recognition of named entities (lecture: 4 hours) 6. Syntactic level: parsing techniques (lecture: 3 hours).



	<p>7. Introduction to word sense disambiguation (lecture: 3 hours). 8. Distributional Semantics (lecture: 3 hours). 9. Laboratory: NLP tools and applications, (NLTK, spaCy, gensim), detection of lexical semantic change, sentiment analysis, implementation of NLP pipelines, transformers, exercises on NLP tasks, e.g. EVALITA (lab: 30 hours).</p>		
<p>Books and bibliography</p>	<p>[JurMa] D. Jurafsky and J. Martin, <i>Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition</i>. Prentice-Hall Series in Artificial Intelligence, 2000. ISBN: 0130950696. [ManSchu] C. Manning and H. Schutze, <i>Foundations of Statistical Natural Language Processing</i>. MIT Press, 2000. [RusNor] S. Russell, P. Norvig, <i>Artificial Intelligence: A Modern Approach</i> (4th ed), Prentice Hall, 2020 (chapters 23-24). [NLP-PY] Steven Bird, Ewan Klein, and Edward Loper. <i>Natural Language Processing with Python</i>. http://www.nltk.org/book (Text is available under the Creative Commons Attribution Noncommercial No-Derivative-Works 3.0 US License). The following section specifies in detail the correspondence between the topics of the program and the chapters of the reference books. Students who wish can obtain texts on loan from the Library. It is suggested to check the possible availability of texts through the University Library System https://opac.uniba.it/easyweb/w8018/index.php? and contact the library to arrange the loan.</p>		
<p>Notes to books and bibliography</p>	<p>For each topic of the program, references to the material and chapters of the texts from which to study are specified below.</p> <ol style="list-style-type: none"> 1. Introduction to NLP and computational linguistics. Lecturers' handout, [RusNor] Chap. 23. 2. Lexical level: tokenization, regular expressions, text normalization. Lecturers' handout, [JurMa] Chap. 2. 3. Language models: n-grams, probability of a sentence, evaluation of language models. Lecturers' handout, [JurMa] Chap. 3, [RusNor] Chap. 23. 4. Classification of documents and applications. Lecturers' handout, [JurMa] Chap. 4. 5. Part-of-Speech (PoS) Tagging: approaches for PoS tagging, recognition of named entities. Lecturers' handout, [JurMa] Chap. 8, [RusNor] Chap. 23. 6. Syntactic level: techniques for parsing. Lecturers' handout, [JurMa] Chap. 17-18, [ManSchu] Chap. 11-12. 7. Introduction to word sense disambiguation. Lecturers' handout. [JurMa] Chap. 23. 8. Distributional semantics. Lecturers' handout. [JurMa] Chap. 6, [RusNor] Chap. 24. 9. Laboratory: NLP tools and applications. Lecture notes and code, [NLP-Py] for programming in Python and using NLTK. <p>Handouts and other material (scientific articles, sample code, etc.) are available on the ADA e-learning platform https://elearning.di.uniba.it</p>		
<p>Teaching schedule and effort</p>			
<p>Hours</p>			
<p>Total</p>	<p>Lectures</p>	<p>Laboratory and exercises</p>	<p>Self-study</p>
<p>150 hours</p>	<p>32 hours</p>	<p>30 hours</p>	<p>88 hours</p>
<p>CFU/ETCS</p>			
<p>6 credits</p>	<p>4 C FU</p>	<p>2 CFU</p>	



Teaching methods	
	<ul style="list-style-type: none">• Lectures, exercises and laboratory sessions carried out with the help of slides made available through the e-learning platform;• During the laboratory sessions, students will practice with the main NLP tools and will develop solutions for some NLP tasks (see section “<i>Learning Assessment Modes</i>” for further details).
Expected learning outcomes	
Knowledge and understanding	<ul style="list-style-type: none">• Fundamentals of natural language processing (NLP);• Theoretical, methodological, and operational aspects of NLP with particular reference to the main levels of linguistic analysis.
Applied knowledge and understanding	<ul style="list-style-type: none">• Apply knowledge on the fundamentals and tools of NLP to address problems related to the management of textual documents;• Exploit NLP techniques for the design of complex systems that handle unstructured data (e.g., text mining applications).
Soft skills	<p>Autonomy of judgment</p> <ul style="list-style-type: none">• Ability to evaluate the correctness of the proposed solution to problems related to the design of AI systems that manage unstructured data (e.g., dialoguing agents, AI systems with NLP interface);• Ability to find, evaluate and adapt known solutions and tools to design NLP pipelines for specific problems (e.g., sentiment analysis or opinion mining). <p>Communication skills</p> <ul style="list-style-type: none">• Ability to communicate - both at a technical level and at the level of scientific dissemination - the choices on the techniques adopted in an NLP pipeline designed to solve a given problem. <p>Ability to learn autonomously</p> <ul style="list-style-type: none">• Ability to follow the advances in the NLP area, learning from both books and scientific articles published in conference proceedings or journals.
Assessment and feedback	
Method of assessment	<p>The exam consists of an oral test, as follows:</p> <ul style="list-style-type: none">• The student will develop a solution for an NLP task of his/her choice, selected from a list proposed by the lectures during the course (e.g. "Hate Speech Detection"). This work can be done in collaboration with another student (not mandatory);• The student will prepare a short technical report (a few pages) and a presentation of the proposed solution. In the case of collaborative work, each student must discuss a specific aspect of the work;• The oral exam consists of a discussion on the proposed solution, as well as on the topics in the program. Duration of discussion: about 20 minutes.



Evaluation criteria	<ul style="list-style-type: none">● Knowledge and understanding:<ul style="list-style-type: none">○ Ability to accurately and clearly describe the techniques and theoretical foundations of NLP;● Applied knowledge and understanding:<ul style="list-style-type: none">○ Ability to adopt the most appropriate NLP techniques for the problem addressed;● Making judgements:<ul style="list-style-type: none">○ Ability to justify the solution adopted among several options;● Communication skills:<ul style="list-style-type: none">○ Clarity of answers/solutions to questions and problems;● Learning skills:<ul style="list-style-type: none">○ Abstraction, reasoning by analogy and demonstration of creativity in problem solving.
Assessment criteria and attribution of the final grade	<p>The final grade is expressed as $x/30$.</p> <p>The final grade takes into account the following aspects:</p> <ol style="list-style-type: none">1) technical correctness of the proposed solution to the task addressed;2) correctness and completeness of the answers to the questions during the oral exam;3) clarity of exposure. <p>To pass the exam, the student must be able to propose a solution that meets at least aspect 1). Ratings above the minimum are assigned when the other aspects are also met.</p> <p>Honors are given when the logic followed by the student highlights particular capacities of abstraction, reasoning by analogy, creativity.</p>
Other	<p>Students are advised to rely exclusively on the information / communications provided on the official websites of the Department of Computer Science, or on social groups only if constituted and administered exclusively by the lecturers of the related courses:</p> <ul style="list-style-type: none">● https://www.uniba.it/it/ricerca/dipartimenti/informatica/en/didactics/degree-courses● https://www.uniba.it/it/ricerca/dipartimenti/informatica/en● https://elearning.di.uniba.it/ <p>The course programs are available here:</p> <ul style="list-style-type: none">● https://programmi.di.uniba.it/ <p>The information that all students should know is written in the Educational Regulations and Study Manifestos available on the site:</p> <ul style="list-style-type: none">● https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea/corsi-di-laurea <p><i>Students are advised to be wary of information and materials circulating on unofficial sites or social groups, as they have proved to be unreliable, incorrect, or incomplete. For any doubt, arrange a meeting with the lecturers according to the procedures provided for the reception.</i></p> <p>Link to the e-learning platform of the ADA department: https://elearning.di.uniba.it</p>



Principali informazioni sull'insegnamento	
Denominazione dell'insegnamento	Natural Language Processing
Corso di studio	Laurea Magistrale in Computer Science (LM-18)
Anno Accademico	2022/23
Crediti formativi universitari (CFU) / European Credit Transfer and Accumulation System (ECTS)	6 CFU
Settore Scientifico Disciplinare	ING-INF/05
Lingua di erogazione	Inglese
Anno di corso	Primo
Periodo di erogazione	2° 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	https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea/computer-science/computer-science

Docente/i	
Nome e cognome	Marco de Gemmis
Indirizzo mail	marco.degemmis@uniba.it
Telefono	080-5443283
Sede	Dipartimento di Informatica, Via Orabona 4, 70125, Bari. Stanza n.761, 7° piano.
Sito web del docente	https://www.uniba.it/it/docenti/degemmis-marco
Nome e cognome	Pierpaolo Basile
Indirizzo mail	pierpaolo.basile@uniba.it
Telefono	080-5442301
Sede	Dipartimento di Informatica, Via Orabona 4, 70125, Bari. Stanza n.758, 7° piano.
Sito web del docente	http://www.di.uniba.it/~swap/index.php?n=Membri.Basile
Ricevimento	appuntamento da concordare per email con i docenti
Sede virtuale	Piattaforma ADA - https://elearning.di.uniba.it

Syllabus	
Obiettivi formativi	<ul style="list-style-type: none">• Conoscenza dei fondamenti teorici dell'elaborazione del linguaggio naturale;• Conoscenza delle problematiche pratiche per lo sviluppo di sistemi di NLP.
Prerequisiti	Le seguenti conoscenze preliminari facilitano la comprensione degli argomenti dell'insegnamento: <ul style="list-style-type: none">• Nozioni di base di: Teoria delle probabilità, programmazione
Contenuti di insegnamento (Programma)	La numerazione degli argomenti ne indica l'ordine di presentazione e di studio. Inoltre, l'indice di ogni argomento è utile per ritrovare i relativi contenuti nei testi di riferimento. <ol style="list-style-type: none">1. Introduzione al NLP ed alla linguistica computazionale (lezione frontale: 4 ore).2. Livello lessicale: tokenizzazione, espressioni regolari, normalizzazione del testo (lezione frontale: 4 ore).3. Modelli linguistici: n-grammi, probabilità di una frase, valutazione dei modelli linguistici (lezione frontale: 6 ore).4. Classificazione di documenti ed applicazioni (lezione frontale: 5 ore).5. Part-of-Speech (PoS) Tagging: approcci per il PoS tagging, riconoscimento di entità con nome (lezione frontale: 4 ore)6. Livello sintattico: tecniche per il parsing (lezione frontale: 3 ore).



	<p>7. Introduzione alla disambiguazione del senso delle parole (lezione frontale: 3 ore).</p> <p>8. Semantica Distribuzionale (lezione frontale: 3 ore).</p> <p>9. Laboratorio: strumenti e applicazioni NLP, (NLTK, spaCy, gensim), rilevamento del cambiamento semantico lessicale, sentiment analysis, implementazione di pipeline di NLP, transformers, esercitazioni su task di NLP, ad esempio EVALITA (ore di esercitazione pratica: 30).</p>		
<p>Testi di riferimento</p>	<p>[JurMa] D. Jurafsky and J. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Prentice-Hall Series in Artificial Intelligence, 2000. ISBN: 0130950696.</p> <p>[ManSchu] C. Manning and H. Schutze, Foundations of Statistical Natural Language Processing. MIT press, 2000.</p> <p>[RusNor] S. Russell, P. Norvig, Artificial Intelligence: A Modern Approach (4th ed), Prentice Hall, 2020 (chapters 23-24).</p> <p>[NLP-Py] Steven Bird, Ewan Klein, and Edward Loper. Natural Language Processing with Python. http://www.nltk.org/book (Il testo è disponibile con licenza Creative Commons Attribution Noncommercial No-Derivative-Works 3.0 US License.)</p> <p>Nella sezione seguente è specificata in dettaglio la corrispondenza tra gli argomenti del programma ed i capitoli dei testi di riferimento.</p> <p>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 Bibliotecario di Ateneo https://opac.uniba.it/easyweb/w8018/index.php? e contattare la biblioteca per concordare il prestito.</p>		
<p>Note ai testi di riferimento</p>	<p>Si specificano di seguito, per ogni argomento del programma, i riferimenti al materiale ed ai capitoli dei testi dai quali studiare.</p> <ol style="list-style-type: none"> 1. Introduzione al NLP ed alla linguistica computazionale. Dispensa dei docenti. [RusNor] Cap. 23. 2. Livello lessicale: tokenizzazione, espressioni regolari, normalizzazione del testo. Dispensa dei docenti, [JurMa] Cap. 2. 3. Modelli linguistici: n-grammi, probabilità di una frase, valutazione dei modelli linguistici. Dispensa dei docenti, [JurMa] Cap. 3, [RusNor] Cap. 23. 4. Classificazione di documenti ed applicazioni. Dispensa dei docenti, [JurMa] Cap. 4. 5. Part-of-Speech (PoS) Tagging: approcci per il PoS tagging, riconoscimento di entità con nome. Dispensa dei docenti, [JurMa] Cap. 8, [RusNor] Cap. 23. 6. Livello sintattico: tecniche per il parsing. Dispensa dei docenti, [JurMa] Cap. 17-18, [ManSchu] Cap. 11-12 7. Introduzione alla disambiguazione del senso delle parole. Dispensa dei docenti. [JurMa] Cap. 23. 8. Semantica Distribuzionale. Dispensa dei docenti. [JurMa] Cap. 6, [RusNor] Cap. 24. 9. Laboratorio: strumenti e applicazioni NLP. Dispense e codice dei docenti, [NLP-Py] per la programmazione in Python e l'uso di NLTK. <p>Le dispense ed altro materiale di approfondimento (articoli scientifici, codice di esempio, etc.), sono disponibili sulla piattaforma di e-learning ADA https://elearning.di.uniba.it</p>		
<p>Organizzazione della didattica</p>			
<p>Ore</p>			
<p>Totali</p>	<p>Didattica frontale</p>	<p>Laboratorio ed esercitazioni</p>	<p>Studio individuale</p>



150 ore	32 ore	30 ore	88 ore
CFU/ETCS			
6 CFU	4 CFU	2 CFU	

Metodi didattici	
	<ul style="list-style-type: none">• Lezioni frontali, esercitazioni e sessioni di laboratorio svolte con l'ausilio di diapositive messe a disposizione attraverso la piattaforma e-learning;• Durante le sessioni di laboratorio, gli studenti si eserciteranno con i principali strumenti di NLP e svilupperanno soluzioni per vari task di NLP (si veda la sezione "Modalità di verifica dell'apprendimento" per ulteriori dettagli).

Risultati di apprendimento previsti	
Conoscenza e capacità di comprensione	<ul style="list-style-type: none">• Fondamenti di elaborazione del linguaggio naturale (NLP);• Aspetti teorici, metodologici e operazionali del NLP con particolare riferimento ai principali livelli di analisi linguistica.
Conoscenza e capacità di comprensione applicate	<ul style="list-style-type: none">• Applicare le conoscenze sui fondamenti e sugli strumenti del NLP per affrontare i problemi relativi alla gestione dei documenti testuali;• Sfruttare le tecniche di NLP per la progettazione di sistemi complessi che gestiscono dati non strutturati (es. applicazioni di text mining).
Competenze trasversali	<p>Autonomia di giudizio</p> <ul style="list-style-type: none">• Capacità di valutare la correttezza della soluzione proposta ai problemi legati alla progettazione di sistemi di IA che gestiscono dati non strutturati (es. agenti dialoganti, sistemi di IA con interfaccia NLP);• Capacità di trovare, valutare e adattare soluzioni e strumenti noti per progettare pipeline di NLP per problemi specifici (ad es. analisi del sentiment o mining di opinioni). <p>Abilità comunicative</p> <ul style="list-style-type: none">• Capacità di comunicare - sia a livello tecnico sia a livello di divulgazione scientifica - le scelte sulle tecniche adottate in una pipeline di NLP progettata per risolvere un determinato problema. <p>Capacità di apprendere in modo autonomo</p> <ul style="list-style-type: none">• Capacità di seguire i progressi nell'area NLP, apprendendo sia da libri sia da articoli scientifici pubblicati in atti di conferenze o riviste.

Valutazione	
Modalità di verifica dell'apprendimento	<p>L'esame consiste in una prova orale, come segue:</p> <ul style="list-style-type: none">• Lo studente svilupperà una soluzione per un task di NLP a sua scelta, selezionato da un elenco proposto dai docenti durante il corso (es. "Hate Speech Detection"). Questo lavoro può essere svolto in collaborazione con un altro studente (non obbligatorio);



	<ul style="list-style-type: none">● Lo studente preparerà un breve rapporto tecnico ed una presentazione della soluzione proposta. In caso di lavoro in collaborazione, ogni studente dovrà discutere un aspetto specifico del lavoro;● La prova orale consiste in una discussione sulla soluzione proposta, nonché sugli argomenti presentati a lezione. Durata della discussione: circa 20 minuti.
Criteria di valutazione	<ul style="list-style-type: none">● Conoscenza e capacità di comprensione:<ul style="list-style-type: none">○ Capacità di descrivere con accuratezza e chiarezza le tecniche e i fondamenti teorici del NLP;● Conoscenza e capacità di comprensione applicate:<ul style="list-style-type: none">○ Capacità di adottare le tecniche di NLP più appropriate per il problema affrontato;● Autonomia di giudizio:<ul style="list-style-type: none">○ Capacità di giustificare la soluzione adottata tra più opzioni;● Abilità comunicative:<ul style="list-style-type: none">○ Chiarezza di risposte/soluzioni a domande/problemi;● Capacità di apprendere:<ul style="list-style-type: none">○ Astrazione, ragionamento per analogia e dimostrazione di creatività nella risoluzione dei problemi.
Criteria di misurazione dell'apprendimento e di attribuzione del voto finale	<p>Il voto finale è espresso in trentesimi.</p> <p>Il voto finale tiene conto dei seguenti aspetti:</p> <ol style="list-style-type: none">1) correttezza tecnica della soluzione proposta al task affrontato;2) correttezza e completezza delle risposte ai quesiti durante la prova orale;3) chiarezza di esposizione. <p>Per superare l'esame, lo studente deve essere in grado di proporre una soluzione che soddisfi almeno l'aspetto 1). Valutazioni superiori al minimo sono assegnate quando sono soddisfatti anche gli altri aspetti.</p> <p>La lode è data quando la logica seguita dallo studente 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">● https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea/corsi-di-laurea● https://www.uniba.it/it/ricerca/dipartimenti/informatica● https://elearning.di.uniba.it/ <p>I programmi degli insegnamenti sono disponibili qui:</p> <ul style="list-style-type: none">● https://programmi.di.uniba.it/ <p>Le informazioni che tutti gli studenti dovrebbero conoscere sono scritte nei Regolamenti didattici e manifesti degli studi disponibili nel sito:</p> <ul style="list-style-type: none">● https://www.uniba.it/it/ricerca/dipartimenti/informatica/didattica/corsi-di-laurea/corsi-di-laurea <p><i>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.</i></p> <p>Link all'insegnamento sulla piattaforma e-learning del dipartimento ADA: https://elearning.di.uniba.it</p>