AI: NATURAL LANGUAGE PROCESSING & SEMANTIC ANALYSIS

Grado en Computación e Inteligencia Artificial / Bachelor in Computer Science and Artificial Intelligence BCSAI SEP-2025 AINLP-CSAI.3.M.A

Area Computer Science

Number of sessions: 30

Academic year: 25-26

Degree course: THIRD

Number of credits: 6.0

Semester: 2º

Category: COMPULSORY

Language: English

Professor: **JUAN JOSÉ MANJARÍN COLON**

E-mail: jjmanjarin@faculty.ie.edu

Juanjo Manjarín received the highest grade for his Ph.D. in Theoretical Physics in String Theory and M-Theory from the Universdad Autónoma de Madrid (UAM), after earning an Advanced Studies Diploma (DEA) in Theoretical Physics from the Universdad Autónoma de Madrid (UAM). His teaching experience includes different international universities where he lectured on Complex Variable Analysis, Theoretical Mechanics or Classical Electrodynamics and in the IE University where he has taught Mathematics, Statistics, Econometrics, Social Media Analytics, Design AI, and Programming in R and Python. He also teaches the course of "Math and Stats for Data Analysis" in the Bootcamp for Data Science in the IE Exponential Learning.

He has published a number of papers on international journals on mathematics and theoretical physics and he was reviewer for Mathematical Reviews from 2003 to 2005. His research interests are Quantum Information and Computing and Network Science together with their applications in Data Science.

He has also corporate experience on different TV and cinema production companies: Gestmusic Endemol, 7 y Acción S.L., Hill Valley S.L., 100 Balas S.L. or Zebra Producciones in TV shows such as “El Hormiguero” or “Esto es vida!”, receiving prizes Ondas and Rose d'Or in 2008 and 2009. He was director, producer and post-producer of different short films and now as director of E8 Producciones is recording a documentary film. He also worked in El Pais in the realization of some divulgative science materials.

Office Hours

Office hours will be on request. Please contact at:

jjmanjarin@faculty.ie.edu

SUBJECT DESCRIPTION

Dive into the world where Natural Language Processing (NLP) stands as a cornerstone of the latest tech revolution! Witness groundbreaking innovations like ChatGPT and Bard, which have captivated the globe with their unparalleled capabilities. But believe it, this is just the starting line. Over the next few years, we're set to see NLP and Text Analysis permeate every sector, reshaping the landscape of data-driven decisions. We're on the cusp of decoding human communication protocols in ways we never thought possible. Now, imagine harnessing this prowess for data science, making sense of vast data lakes, and extracting insights that drive real-world impact. Ready to deep dive? Join this course, and elevate your data science journey by mastering the tools and techniques of NLP!

LEARNING OBJECTIVES

The main topics covered in this course are the following:

- Introduction and History of NLP
- NLP Tools & Resources
- Text Processing & Statistics
- Information Retrieval & Sentiment Analysis
- Machine Learning for Text Analysis
- Word Representations
- Advanced NLP with Neural Networks
- Transformers and Pre-Trained Models
- Linguistic Features in NLP
- Question Answering Systems

CORE TOPICS

The 4 Modules of the course will be:

- Module 1: Foundations of NLP and Historical Overview
- Module 2: Machine Learning and Text Analysis Techniques
- Module 3: Advanced Representations in NLP
- Module 4: Deep Learning for NLP
- Module 5: Conclusion and Integrative Practices

TEACHING METHODOLOGY

IE University teaching method is defined by its collaborative, active, and applied nature. Students actively participate in the whole process to build their knowledge and sharpen their skills. Professor's main role is to lead and guide students to achieve the learning objectives of the course. This is done by engaging in a diverse range of teaching techniques and different types of learning activities such as the following:

- Lecture-based Teaching: The course will mainly be taught through lectures that provide an advanced understanding of C programming, including syntax, control structures, functions,

arrays, pointers, and data structures. The lectures will be interactive and include examples and explanations of programming concepts.

- Hands-on Practice: Students will have the opportunity to practice programming exercises and assignments during the class. This will help them gain hands-on experience in writing and debugging C code using proper programming techniques. Students will be encouraged to work in groups and share their knowledge and experience with their peers.
- Assignments: Regular assignments will be given to students to reinforce the programming concepts learned in class. The assignments will be designed to challenge students to think creatively and solve problems related to computer science and AI. Feedback on the assignments will be provided to help students improve their programming skills.
- Online Resources: Students will be given access to online resources such as tutorials, videos, and programming exercises to help them practice programming outside of class. These resources will also provide additional explanations and examples to supplement the lecture-based teaching.
- Project Based Learning: Students will work on a programming project throughout the course, applying the concepts and techniques they learned in class to solve real-world problems related to computer science and AI. This project will be done in groups and will involve identifying a problem, designing a solution, and implementing it using C programming language. The project will also require students to present their work to the class and receive feedback.
- Peer Learning: Students will be encouraged to work in groups and share their knowledge and experience with their peers. This will foster a collaborative learning environment where students can learn from each other and develop teamwork skills.
- Assessments: Regular assessments such as quizzes, tests, and exams will be conducted to evaluate students' understanding of programming concepts and techniques. These assessments will provide feedback to students and help them identify areas where they need to improve.

Learning Activity	Weighting	Estimated time a student should dedicate to prepare for and participate in
Lectures	26.7 %	40.0 hours
Discussions	10.0 %	15.0 hours
Exercises in class, Asynchronous sessions, Field Work	36.7 %	55.0 hours
Group work	16.7 %	25.0 hours
Individual studying	10.0 %	15.0 hours
TOTAL	100.0 %	150.0 hours

AI POLICY

 # Critical GenAI use is encouraged

In this course, the use of generative artificial intelligence (GenAI) is encouraged, with the goal of developing an informed critical perspective on potential uses and generated outputs.

However, be aware of the limits of GenAI in its current state of development:

·If you provide minimum effort prompts, you will get low quality results. You will need to refine your prompts to get good outcomes. This will take work.

·Don't take ChatGPT's or any GenAI's output at face value. Assume it is wrong unless you either know the answer or can cross-check it with another source. You are responsible for any errors or omissions. You will be able to validate the outputs of GenAI for topics you understand.

·AI is a tool, but one that you need to acknowledge using. Failure to do so is in violation of academic honesty policies. Acknowledging the use of AI will not impact your grade.

Suggested format to acknowledge the use of generative AI tools:

I acknowledge the use of [AI systems link] to [specify how you used generative AI]. The prompts used include [list of prompts]. The output of these prompts was used to [explain how you used the outputs in your work].

If you have chosen not to include any AI generated content in your assignment, the following disclosure is recommended:

No content generated by AI technologies has been used in this assignment.

PROGRAM

Disclaimer: *The following description of the material covered is tentative. While an attempt will be made to cover all listed topics and include other advanced topics that will help students throughout their careers in computer science, the pace of the classes will depend on group performance, which may introduce some variations in the syllabus.*

Pre and post-work for each session:

- Necessary readings will be announced before the session.
- Selected exercises will be indicated at the end of each session.

SESSION 1 (LIVE IN-PERSON)

Module 1: Foundations of NLP and Historical Overview

Session 1: Introduction to the Course

Theory. The inaugural session serves as a gateway into the fascinating world of Natural Language Processing (NLP). Participants will be introduced to the foundational concepts, exploring what NLP encompasses, its vast array of applications, and its pivotal role in bridging the communication gap between humans and machines. As we delve into the mechanisms by which machines interpret and respond to human language, attendees will gain insight into the intricate processes that enable technologies like chatbots, voice assistants, and automated translation.

Practice. To make this introductory session interactive and relatable, students will engage in a thoughtful class discussion. This exercise prompts them to reflect on their daily encounters with NLP, whether knowingly or unknowingly. By listing and sharing these experiences—be it through using voice commands with smartphones, getting product recommendations online, or interacting with customer support chatbots—students will foster an appreciation for the omnipresence and impact of NLP in their daily lives. Additionally, we anticipate that this initial discussion will eventually connect with the final session's discussion, allowing students to gauge and appreciate the depth of their learning journey and perceptible improvement in this area.

Learning Blocks:

- 1st slot: 10 min. Knowing each other.
- 2nd slot: 40 min. Course Introduction.
- 3rd slot: 30 min. Class Discussion.

SESSION 2 (LIVE IN-PERSON)

Session 2: The Dawn of Computational Linguistics

Theory. Tracing the journey of natural language processing, this session delves into the formative phases of computational linguistics. From its embryonic stages to significant landmarks like the creation of ELIZA, one of the first computer programs designed to mimic human conversation, we'll chart the progress and key shifts that have shaped this dynamic field.

Practice. Each student is tasked with researching and selecting a pivotal NLP model or application from history. Their challenge is to distill its complexities and prepare a concise, 5-minute overview video to the class, highlighting the model's unique contributions and its lasting impact on the evolution of computational linguistics.

Learning Blocks:

- 1st slot: 60 min. Lecture.
- 2nd slot: 20 min. Milestones selection for the practice.

SESSION 3 (LIVE IN-PERSON)

Session 3: Fundamental Concepts

Theory. This session dives into the foundational pillars of natural language processing. We'll unpack the mechanics behind perceptrons and Multi-Layer Perceptrons, the basics of n-grams, and the intricacies of the backpropagation algorithm—essential building blocks that underpin much of today's NLP technologies.

Practice. Students will engage with a hands-on exercise. Using provided sentences, they will construct a rudimentary n-gram model. The goal? To harness this model's predictive capabilities and determine the subsequent word in a given sequence.

Learning Blocks:

- 1st slot: 60 min. Lecture.
- 2nd slot: 20 min. Explanation and preparation for the practice.

SESSION 4 (LIVE IN-PERSON)

Session 4: Advanced Concepts in NLP

Theory. In this session, we journey further into the realms of natural language processing, focusing particularly on Vector Space Models (VSMs). Through a meticulous exploration of Latent Semantic Analysis (LSA) and Hyperspace Analogue to Language (HAL), we'll shed light on their foundational principles and wide-ranging applications in the NLP landscape.

Practice. Students will grapple with the challenge of visualizing high-dimensional data inherent to NLP. Leveraging the strengths of Vector Space Models and the dimensionality reduction capabilities of Principal Component Analysis (PCA), the task aims to offer a tangible understanding of these abstract concepts.

Learning Blocks:

- 1st slot: 60 min. Lecture.
- 2nd slot: 20 min. Explanation and preparation for the practice.

SESSION 5 (LIVE IN-PERSON)

Session 5: NLP Libraries and Resources in Python

Theory. This session introduces prominent NLP libraries available in Python, such as NLTK, spaCy, Gensim, and Hugging Face's Transformers. These tools are foundational in the NLP community and serve as the backbone for many text processing and modeling tasks.

Practice. Students will engage in a practical task where they are provided a raw dataset and will utilize one or more of the introduced libraries to clean and preprocess the data, highlighting the utilities and functions each library offers for efficient text processing.

Learning Blocks:

- 1st slot: 40 min. Lecture.
- 2nd slot: 40 min. Explanation and preparation for the practice.

SESSION 6 (LIVE IN-PERSON)

Module 2: Machine Learning and Text Analysis Techniques

Session 6: Regular Expressions in NLP

Theory. This session delves into the world of regular expressions, commonly referred to as "regex." These powerful patterns are integral to text processing in NLP, providing precise and efficient methods for text searching, extraction, and manipulation.

Practice. Students will be tasked with crafting regex patterns tailored to identify and extract specific data from text: phone numbers, email addresses, URLs, and monetary values. This hands-on activity underscores the versatility and precision of regular expressions in real-world scenarios.

Learning Blocks:

- 1st slot: 40 min. Lecture.
- 2nd slot: 40 min. Explanation and preparation for the practice.

SESSION 7 (LIVE IN-PERSON)

Session 7: Text-Based Statistics

Theory. This session navigates the quantitative aspects of NLP, emphasizing how statistics underpin many language processing techniques. Participants will be introduced to key measures, including Term Frequency-Inverse Document Frequency (TF-IDF), lexical density, and word frequency distributions, highlighting their crucial role in comprehending and analyzing text data.

Practice. Engaging in a hands-on activity, students will calculate TF-IDF scores for a set of documents. By applying these scores, they'll determine which document is the most pertinent in response to a given query, illustrating the practical application of this statistical measure in information retrieval.

Learning Blocks:

- 1st slot: 60 min. Lecture.
- 2nd slot: 20 min. Explanation and preparation for the practice.

SESSION 8 (LIVE IN-PERSON)

Session 8: Information Retrieval

Theory. This session offers a concise introduction to Information Retrieval (IR), the science behind fetching relevant data from large repositories. The focus will be on understanding how Vector Space Models (VSMs) play a pivotal role in enhancing the efficiency and accuracy of IR systems.

Practice. Students will be tasked with constructing a rudimentary search engine. Using the principles discussed, this engine will sift through a collection of articles to fetch documents that align closely with user queries, offering a tangible understanding of the theoretical content.

Learning Blocks:

- 1st slot: 60 min. Lecture.
- 2nd slot: 20 min. Explanation and preparation for the practice.

SESSION 9 (LIVE IN-PERSON)

Session 9: Introduction to Naive Bayes

Theory. In this session, we'll delve into the mathematics powering the Naive Bayes algorithm, a fundamental tool in NLP. We'll discuss how this technique efficiently tackles the task of classifying text, helping us categorize chunks of information based on their content.

Practice. Students will engage in a hands-on activity where they use the Naive Bayes method to classify news articles. By the end, each article will be sorted into its rightful category, showcasing the practical utility of this algorithm.

Learning Blocks:

- 1st slot: 40 min. Lecture.
- 2nd slot: 40 min. Explanation and preparation for the practice.

SESSION 10 (LIVE IN-PERSON)

Session 10: Logistic Regression in Text Analysis

Theory. This session will demystify logistic regression and its application in analyzing text. We'll see how this powerful method helps in making decisions based on textual information, especially when we need to classify text into two main categories.

Practice. Students will put their knowledge into action by building a sentiment analysis tool. Using logistic regression, this tool will evaluate movie reviews, determining whether they express a positive or negative sentiment towards the film in question.

Learning Blocks:

- 1st slot: 40 min. Lecture.
- 2nd slot: 40 min. Explanation and preparation for the practice.

SESSION 11 (LIVE IN-PERSON)

Module 3: Advanced Representations in NLP

Session 11: Sentiment Analysis

Theory. In this session, we'll delve into the world of sentiment analysis. By understanding the subtle cues and emotions present in text, we can gauge the general feeling or opinion expressed within the content, be it positive, negative, or neutral.

Practice. Students will embark on a hands-on activity, where they'll gather real-time data from a social network on a specific topic. Using the skills acquired, they'll then analyze and interpret the overall sentiment of the collected posts.

Learning Blocks:

- 1st slot: 40 min. Lecture.
- 2nd slot: 40 min. Explanation and preparation for the practice.

SESSION 12 (LIVE IN-PERSON)

Session 12: Deep Dive into Vector Space Models

Theory. This session narrows in on Vector Space Models (VSMs), a cornerstone in NLP. We'll unpack how they represent textual information and why they're vital for understanding relations between different pieces of text.

Practice. Students will get hands-on by creating a tool that assesses how similar two documents are. Using the principles of cosine similarity and VSMs, they'll determine the closeness of content between various texts.

Learning Blocks:

- 1st slot: 40 min. Lecture.
- 2nd slot: 40 min. Explanation and preparation for the practice.

SESSION 13 (LIVE IN-PERSON)

Session 13: Introduction to Word Embeddings

Theory. This session introduces Word Embeddings, starting with the groundbreaking Word2Vec. We'll trace its evolution and understand how it changed the way machines interpret words, paving the way for more advanced successors.

Practice. Students will dive into a practical task where they'll map and visualize Word2Vec embeddings of words tied to emotions. By plotting them in a 2D space, they can see how these words relate in terms of meaning and context.

Learning Blocks:

- 1st slot: 60 min. Lecture.
- 2nd slot: 20 min. Explanation and preparation for the practice.

SESSION 14 (LIVE IN-PERSON)

Session 14: Feedforward Neural Language Models

Theory. In this session, we'll get ourselves into the realm of neural networks, specifically focusing on their application in language processing. We'll understand how these networks 'think' and process language data.

Practice. Students will create a text classifier using a straightforward neural network. This hands-on activity will offer a tangible grasp of how neural structures can categorize textual information.

Learning Blocks:

- 1st slot: 60 min. Lecture.
- 2nd slot: 20 min. Explanation and preparation for the practice.

SESSION 15 (LIVE IN-PERSON)

Session 15: Part of the Speech (POS) Tagging

Theory. In this session we'll uncover how words in a sentence can be identified as nouns, verbs, adjectives, and more, based on their role and meaning.

Practice. Students will take an article and use POS tagging to pinpoint and list down all the nouns, verbs, and adjectives present in it.

Learning Blocks:

- 1st slot: 50 min. Lecture.
- 2nd slot: 30 min. Explanation and preparation for the practice.

SESSION 16 (LIVE IN-PERSON)

Session 16: Named Entity Recognition (NER)

Theory. In this session, we'll explore how to extract specific pieces of information, like people's names, places, or organizations, from chunks of text. It's like giving a spotlight to certain words that stand out because of their importance.

Practice. Students will analyze a news article, using NER to spot and label all the named entities they find within.

Learning Blocks:

- 1st slot: 50 min. Lecture.
- 2nd slot: 30 min. Explanation and preparation for the practice.

SESSION 17 (LIVE IN-PERSON)

Module 4: Deep Learning for NLP

Session 17: Introduction to RNNs

Theory. In this session, we'll explore Recurrent Neural Networks (RNNs). These are a type of neural network designed specifically for sequential data, understanding patterns over time or through sequences. RNNs have memory components that allow them to remember and use information from previous inputs, making them particularly effective for tasks like language processing.

Practice. For our hands-on activity, we'll use an RNN to predict the upcoming word in a given sentence. This exercise will showcase the RNN's ability to process and predict based on sequential data.

Learning Blocks:

- 1st slot: 60 min. Lecture.
- 2nd slot: 20 min. Explanation and preparation for the practice.

SESSION 18 (LIVE IN-PERSON)

Session 18: Deep Dive into LSTMs

Theory. This session delves into Long Short-Term Memory networks (LSTMs), a special kind of RNN that can learn and remember over long sequences and is less susceptible to the vanishing gradient problem. LSTMs have the unique ability to remember long-term dependencies, which has made them a popular choice in many sequential data processing tasks.

Practice. Much like our last exercise with RNNs, students will work on predicting the next word in a given sentence. However, this time, we'll employ an LSTM. This hands-on activity aims to highlight the improvements LSTMs offer over basic RNNs in capturing long-term dependencies in sequences.

Learning Blocks:

- 1st slot: 60 min. Lecture.
- 2nd slot: 20 min. Explanation and preparation for the practice.

SESSION 19 (LIVE IN-PERSON)

Session 19: Transformers in NLP

Theory. In this session, we turn our attention to Transformers, a groundbreaking architecture that has revolutionized the NLP field. With their self-attention mechanisms, Transformers can capture intricate patterns in data and are behind some of the most advanced language models available. Their flexible design has made them a staple for various NLP tasks.

Practice. To get a feel for their power, students will take on the challenge of fine-tuning a transformer model for a specific text classification assignment. This exercise will provide firsthand experience on how Transformers can be adapted for tailored applications.

Learning Blocks:

- 1st slot: 60 min. Lecture.
- 2nd slot: 20 min. Explanation and preparation for the practice.

SESSION 20 (LIVE IN-PERSON)

Session 20: Pre-Trained Language Models (PLMs)

Theory. This session brings to light the prowess of Pre-Trained Language Models (PLMs) such as BERT and GPT. Recognized for their remarkable ability to understand and generate human-like text, these models serve as cornerstones in the NLP landscape. The concept of transfer learning, where these models are adapted for specific tasks without starting from scratch, will also be highlighted.

Practice. Students will get hands-on experience with BERT, a standout among PLMs. The objective will be to extract sentence embeddings using BERT and subsequently visualize these embeddings. This practical exposure will underline the versatility and utility of such models.

Learning Blocks:

- 1st slot: 60 min. Lecture.
- 2nd slot: 20 min. Explanation and preparation for the practice.

SESSION 21 (LIVE IN-PERSON)

Session 21: Advanced Applications of RNNs and LSTMs

Theory. Delving deeper into the realm of recurrent neural networks, this session will spotlight how RNNs and LSTMs have effectively addressed real-world problems. From time-series prediction to language modeling, their iterative structure has revolutionized numerous domains, making them integral in various applications.

Practice. To solidify understanding, students will embark on the task of designing a mini translator. Leveraging the Seq2Seq LSTM models, the goal will be to translate specific phrases, illustrating the power of these neural structures in language transformation tasks.

Learning Blocks:

- 1st slot: 20 min. Lecture.
- 2nd slot: 60 min. Explanation and preparation for the practice.

SESSION 22 (LIVE IN-PERSON)

Session 22: Advanced Applications of Transformers and PLMs

Theory. This session unravels the transformative power of transformer models, particularly in real-world applications. We'll examine how these architectures, backed by self-attention mechanisms, have transcended traditional boundaries in NLP and have become paramount in addressing intricate tasks.

Practice. To concretize the knowledge, participants will harness the capabilities of GPT models. Their mission: design a chatbot tailored for course content inquiries. Through this exercise, attendees will witness firsthand the prowess of transformers in generating coherent and contextually relevant responses.

Learning Blocks:

- 1st slot: 20 min. Lecture.
- 2nd slot: 60 min. Explanation and preparation for the practice.

SESSION 23 (LIVE IN-PERSON)

Session 23: Introduction to Question Answering (QA)

Theory. Diving into the realm of Question Answering (QA), this session will elucidate the foundational principles behind QA systems. Attendees will discover how these systems are architected to extract precise answers from large volumes of data, shining a light on their significance in today's information-driven world.

Practice. Students will be tasked with building a rudimentary QA system. Their canvas: a fixed dataset, reminiscent of a FAQ page.

Learning Blocks:

- 1st slot: 60 min. Lecture.
- 2nd slot: 20 min. Explanation and preparation for the practice.

SESSION 24 (LIVE IN-PERSON)

Session 24: Advanced QA Techniques and Strategies

Theory. Moving beyond the foundational layer of Question Answering, this session will navigate through the state-of-the-art strategies in QA. As the complexity and demands of information retrieval evolve, so too do the techniques, and attendees will gain insights into the cutting-edge methods employed in modern QA systems.

Practice. Students will further hone their skills by fine-tuning a transformer model for a QA endeavor. The focal point: enhancing the chatbot designed in the previous session, ensuring it excels in addressing course content inquiries with precision and accuracy.

Learning Blocks:

- 1st slot: 30 min. Lecture.
- 2nd slot: 50 min. Explanation and preparation for the practice.

SESSION 25 (LIVE IN-PERSON)

Module 5: Conclusion and Integrative Practices

Session 25: Ethics, Bias, and Real-world Challenges in NLP

Theory. In this session, we'll tackle the critical issues surrounding ethics in NLP. We'll dive deep into the biases that might be present in our models, understand the implications of those biases, and discuss the environmental footprint of training large-scale models. Additionally, we'll touch upon the privacy concerns that arise when dealing with language data.

Practice. Students will be provided with an NLP model or application. Your task will be to critically examine it for potential biases or ethical issues.

Learning Blocks:

- 1st slot: 60 min. Lecture.
- 2nd slot: 20 min. Class Discussion.

SESSION 26 (LIVE IN-PERSON)

Session 26: Cutting-edge Trends and Open Problems in NLP

Theory. Introduction to the latest breakthroughs in NLP, open problems that the research community is addressing, and the future trajectory of NLP.

Practice. Group discussion on how one of the emerging trends can be potentially applied in a real-world scenario. Groups can present a brief proposal of a solution to a current challenge using the latest NLP techniques.

Learning Blocks:

- 1st slot: 40 min. Lecture.
- 2nd slot: 40 min. Class Discussion.

SESSION 27 (LIVE IN-PERSON)

Session 27: Group Presentations

Group project presentations - Session 1 (15 min. for each group). Student presentations on NLP applications they've built or case studies they've researched.

SESSION 28 (LIVE IN-PERSON)

Session 28: Group Presentations

Group project presentations - Session 2 (15 min. for each group). Student presentations on NLP applications they've built or case studies they've researched.

SESSION 29 (LIVE IN-PERSON)

Session 29: Review Session

Recap of the entire course content, clarifying doubts.

SESSION 30 (LIVE IN-PERSON)

Session 30: Final Exam

EVALUATION CRITERIA

Participation (5%)

Active participation in-class activities, discussions, and exercises is an especially important aspect in this course because our focus will be on understanding how the concepts discussed in class can be applied in real-world contexts.

Group Project (30%)

The students will work in groups to delve into one of the exercises proposed during the course. It is expected that the students either carry out an implementation of an NLP application or have conducted an in-depth research on one of the studied tools. This group assignment must be submitted through the Turnitin platform, following the guidelines provided by the professor, and students should prepare a 15-minute presentation to deliver to their classmates in class.

Quizzes (10%)

Students will have to complete a quiz after each of the 9 modules that make up the course. These quizzes must be done before the beginning of the next synchronous session.

Exercises (20%)

The submission of the exercises proposed in class will be evaluated. Not all exercises can be solved in the classroom, so students will have one week from the start of the exercise to submit them. The submission will be done individually through Turnitin.

Final exam (35%)

Students will take a quiz through the blackboard platform in which they will be asked about all the course content. The exam will be composed of multiple choice and open-ended questions.

Late Assignments/Presentation:

Will be penalized 5% per 24-hour period, starting on the day they are due. Only in cases of emergency or illness can changes be made to due dates of assignments or projects. ALL such arrangements are the full responsibility of the student and must be made PRIOR to the due date. Failure to confirm any changes to the due date with the professor prior to the due date will result in a grade of zero.

criteria	percentage	Learning Objectives	Comments
Final Exam	35 %		
Exercises	20 %		
Group Project	30 %		
Quizzes	10 %		

Class Participation	5 %		

RE-SIT / RE-TAKE POLICY

Each student has four chances to pass any given course distributed over two consecutive academic years: ordinary call exams and extraordinary call exams (re-sits) in June/July.

Students who do not comply with the 80% attendance rule during the semester will fail both calls for this Academic Year (ordinary and extraordinary) and have to re-take the course (i.e., re-enroll) in the next Academic Year.

Evaluation criteria:

- Students failing the course in the ordinary call (during the semester) will have to re-sit the exam in June / July (except those not complying with the attendance rule, who will not have that opportunity and must directly re-enroll in the course on the next Academic Year).
- The extraordinary call exams in June / July (re-sits) require your physical presence at the campus you are enrolled in (Segovia or Madrid). There is no possibility to change the date, location or format of any exam, under any circumstances. Dates and location of the June / July re-sit exams will be posted in advance. Please take this into consideration when planning your summer.
- The June / July re-sit exam will consist of a comprehensive exam. Your final grade for the course will depend on the performance in this exam only; continuous evaluation over the semester will not be taken into consideration. Students will have to achieve the minimum passing grade of 5 and can obtain a maximum grade of 8.0 (out of 10.0) – i.e., “notable” in the in the re-sit exam.
- Retakers: Students who failed the subject on a previous Academic Year and are now re-enrolled as re-takers in a course will be needed to check the syllabus of the assigned professor, as well as contact the professor individually, regarding the specific evaluation criteria for them as retakers in the course during that semester (ordinary call of that Academic Year). The maximum grade that may be obtained in the retake exam (3rd call) is 10.0.

After ordinary and extraordinary call exams are graded by the professor, you will have a possibility to attend a review session for that exam and course grade. Please be available to attend the session in order to clarify any concerns you might have regarding your exam. Your professor will inform you about the time and place of the review session. Any grade appeals require that the student attended the review session prior to appealing.

- Students failing more than 18 ECTS credits in the academic year after the June-July re-sits will be asked to leave the Program. Please, make sure to prepare yourself well for the exams in order to pass your failed subjects.
- In case you decide to skip the opportunity to re-sit for an exam during the June / July extraordinary call, you will need to enroll in that course again for the next Academic Year as a re-taker and pay the corresponding extra cost. As you know, students have a total of four allowed calls to pass a given subject or course, in order to remain in the program.

BIBLIOGRAPHY

Compulsory

- Jurafsky, D., & Martin, J. H.. (2008). *Speech and Language Processing: An introduction to speech recognition, computational linguistics.* Prentice Hall. ISBN 9780131873216 (Digital)

- Marina de la Cruz Echeandía, Younis R SH Elhaddad, Suzan Awinat, Alfonso Ortega. (2018). *Handbook of Grammatical Evolution.* Springer. ISBN 9783030087722 (Digital)

Chapter: GE and Semantics. The main goal of this chapter is to explain in a comprehensible way the semantic context in formal language theory. This is necessary to properly understand the attempts to extend Grammatical Evolution (GE) to include semantics. Several approaches from different researchers to handle semantics, both directly and indirectly, will be briefly introduced. Finally, previous works by the authors will be described in depth.

Recommended

- Manning, C., & Schutze, H.. (1999). *Foundations of statistical natural language processing.* ISBN 9780262133609 (Digital)
- Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze. (2012). *Introduction to Information Retrieval.* Cambridge University Press. ISBN 9780511809071 (Digital)
- Bird, S., Klein, E., & Loper, E.. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit.* O'Reilly Media, Inc.. ISBN 9780596516499 (Digital)

BEHAVIOR RULES

Please, check the University's Code of Conduct here. The Program Director may provide further indications.

ATTENDANCE POLICY

Please, check the University's Attendance Policy here. The Program Director may provide further indications.

INSTRUCTIONS FOR IN-PERSON SESSIONS

Student behavior in the classroom is a key factor to guarantee that every student has access to a conducive learning environment.

For this reason, IE University expects students to comply with the following rules:

- Punctuality: Students will not be allowed to enter the classroom once the class has begun, nor will they be allowed to leave the classroom until the end of the class. If a student arrives late or leaves early, the professor will mark them absent.
- Disruptive Behavior: The professor may ask students to leave the classroom if they engage in disruptive behavior such as side talking, misuse of digital devices (non-class-related use), eating/drinking, and if they are disrespectful when their peers or the professor are presenting. If this occurs, the professor will mark them absent.
- Smoking: We remind you that smoking or vaping inside the building is not permitted.

- Classroom Hygiene: As a sign of respect to the next students who will use the classroom after the session, students are expected to keep the classroom space clean and tidy and use the trash bins located outside the classroom.

Please refer to the Code of Conduct for further details.

INSTRUCTIONS FOR LIVE ONLINE SESSIONS

Student behavior during online sessions must comply with IE University standards on education, respect for peers and professors, and a commitment to learning. Any infraction to these standards could be considered an ethics violation as per the IE Code of Ethics.

Students must:

- Keep their cameras on for the entire session.
- Be at a desk seated in an appropriate position, in a distraction-free, professional environment, and follow an adequate dress-code for attending classes.
- Be courteous and conduct themselves maturely with their peers and professor in the digital or hybrid environment.
- Use the correct equipment.
- Download and log into (with IE University email) the latest version of Zoom if their course is taught through this platform.

Punctuality is expected from all students. For that reason, students will not be allowed to access the session once the class has begun. If a student is late, leaves the online session before the class ends, or unjustifiably switches off the camera during the session for an extended period of time, they will be marked absent.

RECORDINGS

Recordings of in-person sessions will not be made available. Only in very extraordinary circumstances Program Management can grant an exception to this policy based on justified and documented reasons. In any case, recordings will only be available for a limited period of 80 days.

ETHICAL POLICY

Please, check the University's Ethics Code here. The Program Director may provide further indications.

Student Privacy Statement:

At times, students may disclose personal information through class discussions. It is expected that all members of the class will respect the privacy of their classmates. This means that the information disclosed in the class will not be repeated or discussed with other students outside of the course.

Decisions about Grades:

Decisions about grades are made very carefully, and are final at the end of the course. If you have questions regarding a certain grade or you would like to receive personal feedback, you must request a meeting with me to discuss grades on specific assignments before the last class of the course. Any disputes regarding grades must be resolved before the final exam. “Extra credit” or makeup assignments will only be allowed under extenuating circumstances at the sole discretion of the course professor.

ACADEMIC INTEGRITY

Unless you are specifically instructed to work with other students in a group, all of your assignments, papers, projects, presentations, and any work I assign must reflect your own work and thinking.

What is academic integrity? When you do the right thing even though no one is watching. The core values of integrity, both academic and otherwise include: honesty, fairness, respect, responsibility, and trust. Academic Integrity requires that all students within Instituto de Empresa (IE) act in accordance with these values in the conduct of their academic work, and that they follow the rules and regulations concerning the accepted conduct, practices and procedures of academic research and writing. Academic Integrity violations are defined as Cheating, Plagiarism or other violations of academic ethics.

Cheating and plagiarism are very serious offenses governed by the IE student code of conduct. Any student found cheating or plagiarizing on any assignment or component of this course will at a minimum receive a "0" on the affected assignment. Moreover, the student will also be referred to the University Judicial System for further action. Additional penalties could include a note on your transcript, failing the class, or expulsion from the university.

It is important to note that, while the list below is comprehensive, it should not be considered exhaustive.

Cheating includes:

- An act or attempt to give, receive, share, or utilize unauthorized information or unauthorized assistance at any time for assignments, papers, projects, presentations, tests or examinations.
- Students are permitted to mentor and/or assist other students with assignments by providing insight and/or advice. However, students must not allow other students to copy their work, nor will students be permitted to copy the work of other students. Students must acknowledge when they have received assistance from others.
- Failure to follow rules on assignments, papers, projects, presentations, tests or examinations as provided by the course professor and/or as stipulated by IE.
- Unauthorized co-operation or collaboration.
- Tampering with official documents, including electronic records.
- The impersonation of a student on presentations, exercises, tests or an examination. This includes logging onto any electronic course management tool or program (e.g. Black Board, etc.) using someone else's login and password.

Plagiarism includes:

- Using the work of others and attempting to present it as your own. For example, using phrases or passages from books, articles, newspapers, or the internet and not referencing them properly in your document. This includes using information from others without citing it, misrepresentation of cited work, and misuse of quotation marks.
- Submitting an assignment or paper that is highly similar to what someone else has written (i.e., minimal changes in wording, or where the sentences are similar, but in a different order).
- You don't have to commit "word for word" copying to plagiarize – you can also plagiarize if you turn in something that is "thought for thought" the same as someone else.

Other violations of academic ethics include:

- Not acknowledging that your work or any part thereof has been submitted for credit elsewhere.
- Misleading or false statements regarding work completed.
- Knowingly aiding or abetting anyone in committing any form of an Academic Integrity violation.