COURSE DESCRIPTION CARD - SYLLABUS

Course name
Natural language processing

Course

Field of study
Artificial Intelligence
Area of study (specialization)

Level of study
First-cycle studies
Form of study
full-time

Year/Semester
3/6
Profile of study
general academic
Course offered in
English
Requirements
compulsory

Number of hours

Lecture
30
Tutorials

Laboratory classes
30
Projects/seminars

Other (e.g. online)

Number of credit points

4

Lecturers

Responsible for the course/lecturer:
Mikołaj Morzy, PhD, DSc

email: mikolaj.morzy@put.poznan.pl

Responsible for the course/lecturer:
Agnieszka Ławrynowicz, PhD, DSc

email: alawrynowicz@cs.put.poznan.pl

Dawid Wiśniewski, MSc

email: dawid.wisniewski@cs.put.poznan.pl

Prerequisites

A student beginning this course should have basic knowledge of programming, fundamentals of logic,

fundamentals of artificial intelligence and statistics and data analysis. The student should have the ability to solve basic problems in the field of implementation and cost evaluation of simple algorithms, and the ability to obtain information from indicated sources. The student should also understand the necessity of extending their competences and be willing to cooperate within a team. Moreover, in terms of social competence, a student should demonstrate such attitudes as honesty, responsibility, perseverance, cognitive curiosity, creativity, personal culture, respect for other people.

Course objective

1. To provide students with basic knowledge of natural language processing.

2. To develop in students the ability to solve simple problems in how to use and implement methods and systems using natural language processing.

3. To develop students' skills in applying artificial intelligence and data analysis methods to natural language processing.

4. To develop in students the skills of teamwork.

Course-related learning outcomes

Knowledge

1. The student has a structured and theoretically grounded general knowledge of the key issues of natural language processing and a detailed knowledge of selected topics in this area of science.

2. The student is familiar with the main directions of development and the most important achievements in the field of natural language processing and other related areas such as the use of machine learning methods in text processing and other issues of artificial intelligence related to natural language processing.

3. The student has basic knowledge about the important directions of development and the most important achievements of artificial intelligence (in particular natural language processing) understood as an important field of computer science which draws on achievements of other scientific disciplines and provides them with solutions of practical potential.

Skills

1. The student is able to acquire information from various sources, including literature and databases, both in Polish and English, integrate them properly, interpret and critically evaluate them, draw conclusions and thoroughly justify opinions formulated by them.

2. The student is able to use information and communication technology in various stages of implementation of IT projects.

3. The student is able to formulate and solve artificial intelligence tasks using adequate methods, including analytical, simulation or experimental methods.

4. The student is able to design and implement a device or a widely understood information system according to given specifications, selecting the appropriate programming language for the given programming task and using adequate artificial intelligence methods, techniques and tools.

5. The student is able to formulate algorithms and implement them using at least one popular tool.

6. The student is able to acquire, analyse and process text data sets, protect them against unauthorised access and synthesise them into knowledge and conclusions useful in solving a wide range of problems arising in the work of computer scientists, specialists in the field of artificial intelligence, including problems of industrial, business and administration specificity

7. The student is able to use and adapt models of intelligent behaviors and IT tools where natural language processing is used that simulate these behaviors

8. The student is able to use information and communication techniques, especially those using natural language processing components, and tools at various stages of IT projects, including preparing a well-documented study of a problem, giving an oral presentation, communicating using specialist terminology and discussing various opinions and positions, also in a non-specialist environment

Social competences

1. The student understands that in artificial intelligence knowledge and skills become obsolete very quickly, recognizing the need for continuous training and improving own competence.

2. The student is aware of the significance of knowledge and scientific research connected with computer science and artificial intelligence (in particular natural language processing) in solving practical problems of key importance for the functioning of individuals, companies, organizations and the whole society

3. The student is able to think and act in an entrepreneurial way, i.e. finding commercial applications for created artificial intelligence systems (involving natural language processing), taking into account not only economic benefits, but also legal and social aspects

Methods for verifying learning outcomes and assessment criteria

Learning outcomes presented above are verified as follows:

Formative assessment:

(a) in terms of lectures: on the basis of answers to questions on the material discussed in previous lectures (in the form of tests) and activity.

b) for laboratories: on the basis of the assessment of the current progress of the partial tasks.

Summative assessment:

a) in the scope of lectures, verification of the assumed educational effects is realized by: evaluation of knowledge and skills demonstrated in the presentation resulting from the realization of a project related to natural language processing and summary of points and discussion of tests from the lecture. The final

evaluation of the lectures consists of: points from the tests available after selected lectures, points from the presentation and points for activity.

b) in the scope of laboratory the verification of the assumed learning outcomes is realized by: evaluation of skills related to the implementation of laboratory exercises, evaluation of the performance of tasks implemented partly during the laboratories and partly after their completion, evaluation of the project that students will implement to summarize the acquired knowledge and skills.

Obtaining additional points for the activity during the classes, especially for: the discussion of additional aspects of the problem, the demonstration of interesting skills beyond the program, the efficiency of applying the acquired knowledge when solving the problem.

Pass mark: 50% of the points.

Programme content

1) Fundamentals of Computational Linguistics

introduction to NLP, general overview of the subject

morphology, syntax, sentence dissection, parts of speech and parts of sentence, POS tagging, dependency parsing

sparse vector representations (BOW, BONG, TFIDF)

language modeling by Markov chains, CRFs

2) Dense vector representations

encoder-decoder and word2vec model

static embeddings: GloVe, FastText, Pointcare, dependency-based

3) Information Extraction (NER)

4) Language models

first approaches: RNN, LSTM, GRU

Transformer architecture

BERTology: BERT, CamemBERT, RoBERT

advanced models: Siamese networks, Sentence Transformers, Longformers

5) Applications

text annotation for NLP

datasets, benchmarks, tests

fine-tuning and transfer learning in NLP

Teaching methods

lecture: multimedia presentation, demonstration of exemplary solutions, quizes

laboratory classes: practical exercises, discussion, teamwork, analysis of multimedia materials

Bibliography

Basic

1. Speech and Language Processing (3rd ed. draft), Dan Jurafsky and James H. Martin. Draft chapters in progress, Dec 29, 2021, https://web.stanford.edu/~jurafsky/slp3/

2. Natural Language Processing with Python, Steven Bird, Ewan Klein, and Edward Loper, O'Reilly Media, 2009, http://www.nltk.org/book/

Additional

1. Natural Language Processing in Action. Understanding, analyzing, and generating text with Python, Hobson Lane, Cole Howard, Hannes Hapke, Manning Publications, 2019

2. Practical Natural Language Processing: A Pragmatic Approach to Processing and Analyzing Language Data: A Comprehensive Guide to Building Real-World NLP Systems, Sowmya Vajjala, Bodhisattwa Majumder i in., O'Reilly Media, 2020

3. Natural Language Processing with PyTorch: Build Intelligent Language Applications Using Deep Learning 1st Edition, Delip Rao, Brian McMahan, O'Reilly Media, 2019

4. Foundations of Statistical Natural Language Processing, Chris Manning and Hinrich Schütze, MIT Press. Cambridge, MA: May 1999, http://nlp.stanford.edu/fsnlp/

5. The Text Mining Handbook, Ronen Feldman, James Sanger, Cambridge University Press, 2007

6. Rothman, Denis. Transformers for Natural Language Processing: Build innovative deep neural network architectures for NLP with Python, PyTorch, TensorFlow, BERT, RoBERTa, and more. Packt Publishing Ltd, 2021.

7. Ravichandiran, Sudharsan. Getting Started with Google BERT: Build and train state-of-the- art natural language processing models using BERT. Packt Publishing Ltd, 2021.

8. Altinok, Duygu. Mastering spaCy: An end-to-end practical guide to implementing NLP applications using the Python ecosystem. Packt Publishing Ltd, 2021.

Breakdown of average student's workload

	Hours	ECTS
Total workload	105	4,0
Classes requiring direct contact with the teacher	60	2,5
Student's own work (literature studies, preparation for laboratory classes, preparation for tests, project preparation) [1]	45	1,5

[1] delete or add other activities as appropriate