COURSE DESCRIPTION CARD - SYLLABUS

Course name
Selected Topics in Natural Language Processing [S2SI1E>PJN]

Course

Field of study
Artificial Intelligence

Year/Semester
1/2

Area of study (specialization)
–

Profile of study
general academic

Level of study
second-cycle

Course offered in
English

Form of study
full-time

Requirements
compulsory

Number of hours

Lecture
0

Laboratory classes
0

Other
0

Tutorials
0

Projects/seminars
30

Number of credit points

2,00

Coordinators

dr inż. Mateusz Lango
mateusz.lango@put.poznan.pl

Lecturers

Prerequisites

A student starting this course should have basic knowledge of probability and statistics (normal, binomial, Dirichlet, Beta and Bernoulli distributions, maximum likelihood estimation, KL divergence, Monte Carlo methods, statistical hypothesis testing), in-depth knowledge of machine learning (ensembles, k-NN, Naive Bayes, SVM, curse of dimensionality) and deep learning (multi-layer neural networks, recurrent encoder-decoder networks, convolutional networks, transformer, backpropagation, GAN, VAE, triplet loss). Additionally, intermediate knowledge of text processing, equivalent to the course "Natural language processing," is also assumed (regular expressions, stemming, lemmatization, stopwords, bag-of-words model, measures of text similarity, BERT/GPT-x/T5 models, CRF, MEMM, dependency parsing, constituency parsing, IBM translation models, WordNet etc.). The student should have the ability to obtain information from indicated sources, read and understand a scientific paper as well as present the research to the public in a clear way. In terms of social competencies, the student must understand that in computer science knowledge and skills quickly become obsolete. The student should present attitudes such as honesty, responsibility, perseverance, cognitive curiosity, creativity, and respect for other peoples.

Course objective

The aim of the course is to familiarize students with the methodology and recent research advances in natural language processing and human language technologies. Classes focus on the discussion of recent academic papers published in the top venues such as ACL, NAACL, COLING, EMNLP that concern problems such as automatic translation, sentiment analysis, dialogue systems, natural language generation, question answering, syntax analysis, and topic modeling. The additional goal of the course is to develop the ability to critically analyze results of statistical and machine learning models in various respects (computational complexity, type of training data and required sample size, model assumptions / limitations, inference methods).

Course-related learning outcomes

Knowledge 1. Student has advanced and in-depth knowledge of the construction of computer systems that process natural language using statistical methods - [K2st_W3]
2. Student has an in-depth understanding of the architectures of deep neural networks used in human language technologies (in particular recursive and recursive architectures) - [K2st_W3]
3. Student has advanced and in-depth knowledge related to selected issues, such as language modeling, syntax analysis, distributional semantics, named-entity recognition, machine translation - [K2st_W3]
4. Student knows development trends and the essential new achievements of linguistic engineering (including modern deep machine learning architectures) – [K2st_W4]
5. Student knows advanced methods, techniques, and tools used in the construction of dialogue systems, translators, parsers, and question answering systems - [K2st_W6]
6. Student understands advanced methods used in research in the field of linguistic engineering - [K2st_W6]
Skills 1. Student is able to obtain information on linguistic engineering techniques from literature and other sources, integrate them, interpret and critically evaluate them, draw conclusions and justify opinions - [K2st_U1]
2. Student can assess the usefulness and the possibility of using new achievements of machine learning to solve problems in linguistic engineering - [K2st_U6]
3. Student can determine the directions of a further self-study - in particular for learning new techniques of state-of-the-art linguistic engineering [K2st_U16]
Social competences 1. Student understands that in human language technologies, knowledge and skills become obsolete very quickly - [K2st_K1]
2. Student understands the importance of using the latest achievements in the field of human language technologies and machine learning in solving practical problems - [K2st_K2]

Methods for verifying learning outcomes and assessment criteria

Learning outcomes presented above are verified as follows:

The knowledge presented during the seminar will be verified by a written test containing open-ended and multiple-choice questions, as well as by the assessment of the presentation prepared by the student based on a selected NLP research paper. Additionally, the progress of the student will be continuously evaluated by in-class discussions of NLP methods/problems/approaches.

The following grading scale is used: above 51% of points - satisfactory (3.0), 61% - satisfactory plus (3.5), 71% - good (4.0), 81% - good plus (4.5), 91% - very good (5.0).

Programme content

During the seminar, students will learn about current issues and research trends in NLP, as well as modern applications of NLP technology.

Course topics

The course introduces the student to conducting research in the field of natural language processing, along with its many tasks and research best practices such as ensuring reproducibility of scientific results. During the seminar, students will learn about current research problems and trends in NLP, as well as modern applications of NLP technology, such as machine translation, sentiment analysis, text-to-speech, speech recognition (ASR), question answering, information retrieval, dialogue systems and others. Students will also learn how to pre-assess the quality of presented scientific research, how to evaluate the relevance of a given scientific paper to the problem being solved, and become familiar with the structure and language

used in scientific papers. The seminar also aims to raise awareness of the growing role that NLP technologies play in modern society and related issues of professional ethics.

Teaching methods

1. Seminar: presentation, discussion of issues and solutions

Bibliography

1. Selected academic papers from top-tier NLP venues (ACL, EACL, EMNLP, NAACL,...)
2. Jurafsky D., Martin J.H.: Speech and Language Processing, III edycja, Pearson/Prentice Hall, 2022 (dostęp online: https://web.stanford.edu/~jurafsky/slp3/)
3. Li Deng, Yang Liu: Deep Learning in Natural Language Processing. Springer, 2018 (access through eZasoby service of PUT library)
Additional
1. Goodfellow I., Yoshua B., Courville A.: Deep Learning. MIT Press, 2016
2. Goldberg, Y., Neural Network Methods in Natural Language Processing, Springer, 2017

Breakdown of average student's workload

	Hours	ECTS
Total workload	50	2,00
Classes requiring direct contact with the teacher	30	1,50
Student's own work (literature studies, preparation for laboratory classes/ tutorials, preparation for tests/exam, project preparation)	20	0,50