COURSE DESCRIPTION CARD - SYLLABUS

Course name
Natural language processing [S1Inf1>PJN]

Course

Field of study
Computing

Year/Semester
3/6

Area of study (specialization)
–

Profile of study
general academic

Level of study
first-cycle

Course offered in
Polish

Form of study
full-time

Requirements
elective

Number of hours

Lecture
24

Laboratory classes
20

Other
0

Tutorials
0

Projects/seminars
0

Number of credit points

2,00

Coordinators

dr inż. Dawid Wiśniewski
dawid.wisniewski@put.poznan.pl

dr hab. inż. Agnieszka Ławrynowicz prof. PP
agnieszka.lawrynowicz@put.poznan.pl

Lecturers

Prerequisites

A student starting this subject should have basic knowledge in programming, basics of logic, basics of artificial intelligence as well as statistics and data analysis. The student should have the ability to solve basic problems in the field of implementation and assessment of the operation cost of simple algorithms and the ability to obtain information from the indicated sources. The student should also understand the need to expand his/her competences and demonstrate readiness to cooperate as part of the team. Moreover, in terms of social competences the student must present attitudes such as honesty, responsibility, perseverance, curiosity, creativity, personal culture, and respect for other people.

Course objective

1. Providing students with basic knowledge on natural language processing. 2. Developing students" ability to solve simple problems in terms of ways of the use and implementation of methods and systems that use natural language processing. 3. Developing students" ability to apply artificial intelligence methods and data analysis to natural language processing. 4. Shaping students" teamwork skills.

Course-related learning outcomes

Knowledge:
A student:
1. Has a structured and theoretically based general knowledge of key issues regarding natural language processing, and detailed knowledge of selected areas of this scientific field.
2. Has knowledge of important 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 for word processing and other areas of artificial intelligence related to natural language processing.
3. Knows the basic techniques, methods and tools used in the process of solving tasks related to natural language processing, mainly of an engineering nature, in the field of key IT issues.

Skills:
A student:
1. Can obtain information from various sources, including literature and databases, both in Polish and in English, integrate them properly, interpret them and critically evaluate them, draw conclusions and exhaustively justify the opinions he/she formulates.
2. Can properly use information and communication techniques applied at various stages of the implementation of IT projects.
3. Is able to formulate and solve IT tasks, apply appropriately selected methods including analytical, simulation or experimental methods.
4. Can - according to the given specification - design and implement the device or widely understood IT system, selecting the appropriate programming language for a given programming task and using appropriate methods, techniques and tools.
5. Has the ability to formulate algorithms and their implementation using at least one of the popular tools.
6. Can plan and implement the process of own permanent learning and knows the possibilities of the further training (2nd and 3rd cycle studies, postgraduate studies, courses and examinations conducted by universities, companies and professional organizations).

Social competences:
A student:
1. understands that knowledge and skills very quickly become obsolete in computer science.
2. Is aware of the importance of knowledge in solving engineering problems and knows examples and understands the causes of the malfunctioning information systems that led to serious loss of health and even life.

Methods for verifying learning outcomes and assessment criteria

Learning outcomes presented above are verified as follows:

Learning outcomes presented above are verified as follows:
Formative assessment:
a) lectures: on the basis of answers to questions about the material discussed in previous lectures (in the form of quizes) and activities.
b) laboratories: based on the evaluation of the current progress in the implementation of sub-tasks.
Summative assessment:
a) lectures: verification of the assumed learning outcomes is carried out by evaluation of knowledge and skills demonstrated in the analysis and presentation resulting from the indicated problem of Natural Language Processing and summarizing points and discussion of the quizes. The final grade for the lectures consists of: points from the quizes available after selected lectures, points from the presentation and points for activity.
b) laboratories: verification of the assumed learning outcomes is carried out by skills assessment related to the implementation of laboratory exercises, assessment of the implementation of tasks partly during the laboratories and partly after their completion, assessment of the project that the students carry out in order to consolidate the acquired knowledge and skills. Obtaining additional points for activity during classes, especially for: discussing additional aspects of the issue, demonstrating interesting extracurricular skills, the effectiveness of applying the acquired knowledge when solving a given problem, remarks leading to the improvement of teaching materials or the teaching process. Passing threshold: 50% of points.

Programme content

The lecture covers the following topics:
1. Basic architecture of natural language processing systems.
2. Basic text processing techniques (regular expressions, filtering function words, segmentation, lemmatization, editing distance).
3. Text classification.
4. Grammatical analysis (recognition of parts of speech, tagging of parts of speech, dependency parsers).
5. Statistical language modeling, distributed representations and embedding vectors: n-gram, word2vec, GloVe, softmax.
6. Representation of linguistic knowledge in the form of thesauri, dictionaries (WordNet, FrameNet).
7. Application of deep learning to natural language processing (GRU, LSTM, other models, e.g. recursive networks).
8. Information and knowledge extraction from text (detection of named entities, extraction of relations).
9. Sentiment and opinion analysis.
10. Dialogue systems, chatbots.
The program of laboratories covers the basic methods of natural language processing, using popular libraries and toolkits (e.g. Python NLTK library, SpaCy, gensim, BeautifulSoup), applying deep learning to natural language processing (using, for example, TensorFlow / Keras / PyTorch) and case studies in the field of applications of known methods (analysis of sentiment and opinion, automatic summarization), in particular:
1. Fast text search / manipulation - regular expressions.
2. Classification of texts (creating a BagOfWords representation, Tokenization, Lematisation, Stemming, TF-IDF normalization, SVM, Naive Bayes).
3. NGrams (NGram vs BagOfWords representation, language detection, text generation).
4. Embeddings as a low-dimensional alternative to BagOfWords / NGram (similarity in the space of embedding, embeddings for classification).
5. Correcting typos using editing distance.
6. Neural networks (networks as a sequence of operations on matrices, implementing a network without using frameworks).
7. Resource creation (Crawling / Scraping of data from the web, creating resources).
8. Recursive networks in text processing (RNN from scratch without the use of frameworks, the idea of memory in RNN).
9. Sentiment detection using advanced network architectures (GRU / LSTM / CNN).
10. Named entities and noun phrases detection.
11. Summarization through Key-sentence extraction, topic modeling (LDA).
12. Information extraction (dependency tree, grammatical sentence breakdown).

Course topics

none

Teaching methods

lecture: multimedia presentation, demonstration of sample solutions, quizes
laboratories: practical exercises, discussion, team work, analysis of materials in the form of multimedia

Bibliography

Basic
1. Natural Language Processing with Python, Steven Bird, Ewan Klein, and Edward Loper, O"Reilly Media, 2009, available online http://www.nltk.org/book/
Additional
1. Speech and Language Processing (3rd ed. draft), Dan Jurafsky and James H. Martin. Draft chapters in progress, October 16, 2019
2. Natural Language Processing in Action. Understanding, analyzing, and generating text with Python, Hobson Lane, Cole Howard, Hannes Hapke, Manning Publications, 2019
3. Foundations of Statistical Natural Language Processing, Chris Manning and Hinrich Schütze, MIT Press. Cambridge, MA: May 1999, http://nlp.stanford.edu/fsnlp/
4. The Text Mining Handbook, Ronen Feldman, James Sanger, Cambridge University Press, 2007

5. Inżynieria lingwistyczna. Komputerowe przetwarzanie tekstów w języku naturalnym, Agnieszka Mykowiecka, 2007, Wydawnictwo PJWSTK, Seria: Podręczniki akademickie.
6. Ian Goodfellow, Yoshua Bengio, Aaron Courville "Deep Learning" MIT Press 2016, http://www.deeplearningbook.org
7. Zaprojektuj bota. Tworzenie interfejsów konwersacyjnych, Shevat Amir, Helion, 2018

Breakdown of average student's workload

	Hours	ECTS
Total workload	60	2,00
Classes requiring direct contact with the teacher	46	1,50
Student's own work (literature studies, preparation for laboratory classes/ tutorials, preparation for tests/exam, project preparation)	14	0,50