STUDY COURSE DESCRIPTION FORM

Name of the course	Code
Machine learning methods in natural language processing	
Name of the doctoral school	Year /Semester
Poznan University of Technology Doctoral School	**….**
Specialty	Type (obligatory, elective):
Information and communication technology	**elective**
No. of hours	No. of credits
Lectures: **4** Classes: **-** Laboratories: **-** Seminars: **-**	**1**

Cycle of study:	**Form of study:**	**Assessment:** (written exam, presentation, etc.)
Third-cycle studies (Polish Qualifications Framework level eight)	Full-time	short presentation

Responsible for the course/lecturer:

dr hab. inż. Mikolaj Morzy, prof. PUT
e-mail: Mikolaj.Morzy@put.poznan.pl
phone : +48 61 665 2961
Faculty of Computing and Telecommunications
Poznan University of Technology
Piotrowo street 2, 60-965 Poznan, Poland

Prerequisites in terms of knowledge, skills and social competencies:

1	**Knowledge:** very basic understanding of algorithms, ability to read and understand very simple blocks of computer pseudo-code, understanding of basic statistics
2	**Skills:** ability to transfer knowledge between domains, and to apply learned patterns to different domains
3	**Social competencies:** thinking outside the box to solve various problems using available textual resources.

Objectives of the course:

The main aim of the course is to present the newest developments in the area of natural language processing (NLP) using algorithms and techniques of machine learning (ML). The majority of human knowledge is currently stored in the form of unstructured text. Abstracts, reviews, descriptions, posts, emails, tweets, all create a huge corpus of data which cannot be analyzed manually. Such textual corpora exist in almost all domains of science and technology. Computer methods for text analysis are collectively known as NLP. In the recent years we are witnessing a true revolution in NLP due to the development of machine learning methods designed specifically to tackle NLP challenges. During the lecture the students will learn basic NLP methods (tokenization, lemmatization, stemming), basic representation methods (one-hot encoding, TF-IDF), as well as methods based on neural networks (word and sentence vectors, transformer language models). We will discuss methods for sentiment analysis in text, named entity recognition, neural translation, sequence to sequence learning, and more.

Educational results (Study outcomes)

Knowledge:

P8S_WG	Students understand current achievements in the combined fields of machine learning and natural language processing, they understand basic principles of algorithms used to extract useful knowledge from unstructured text.	**SzD_W01**

P8S_WG	Students know the current developmental trends in machine learning and natural language processing, and can identify research questions in their scientific domains that can be addressed using machine learning and natural language processing.	**SzD_W02**
Skills:		
P8S_UW	Students use the knowledge of machine learning and natural language processing to collect new data and new insights in their respective scientific disciplines	**SzD_U01**
P8S_UW	Students can design new distributed representations of data in their scientific disciplines using the paradigm of encoder-decoder neural network architecture	**SzD_U03**
Social competencies:		
P8S_KK	Students acknowledge the importance of natural language processing methods by designing a research question involving their own discipline that can be addressed using machine learning and natural language processing.	**SzD_W03**

Compulsory literature:

1. Collobert, Ronan, et al. "*Natural language processing (almost) from scratch*." Journal of machine learning research 12. Aug (2011): 2493-2537.

Additional literature:

1. Mikolov, Tomas, et al. "*Distributed representations of words and phrases and their compositionality*." Advances in neural information processing systems. 2013.
2. Le, Quoc, and Tomas Mikolov. "*Distributed representations of sentences and documents*." International conference on machine learning. 2014.
3. Bengio, Yoshua, et al. "*A neural probabilistic language model*." Journal of machine learning research 3.Feb (2003): 1137-1155.

COURSE DESCRIPTION

	General issues	Specific issues	No. of hours
1	Natural Language Processing	• Representation of text • Stemming, lemmatization, tokenization • Vector-space models of text	1
2	Machine learning	• Introduction to machine learning • Neural networks • Classification and clustering • Distance functions	1
3	Distributed representations	• Word vectors (word2vec, GloVe) • Sentence vectors • Language models	1
4	Advanced topics	• Sentiment analysis • Named Entity Recognition • Neural Translation	1

Assessment methods of educational results

Students will be asked to prepare a short (5-10 minute) presentation which will outline a plan to employ methods presented during the lecture to address a research question in their scientific discipline. Each presentation will be discussed among all participating students.

STUDENT'S WORKLOAD	
Activity	**Hours**
Participation in lectures, classes, seminars and laboratories	4
Contact hours with lecturers	2
Self-study	10
Preparation of presentation	10
TOTAL	**26**
TOTAL NUMBER OF ECTS POINTS FOR THE COURSE	1