DOCTORAL (PHD) STUDIES

COURSE UNIT DESCRIPTION

Course unit title	Scientific areas	Faculty	Institute, department
Natural Language Processing	Informatics Engineering (T 007)	Faculty of Mathematics and Informatics	Institute of Data Science and Digital Technologies

Study method	Number of credits	Study method	Number of credits
Lectures	1 (spring semester)	Consultations	1
Individual works	4	Seminars	1

Summary

It is desirable for a doctoral student to know linear algebra, mathematical analysis, and Python programming language.

The aim of this subject is to improve knowledge of natural language processing using advanced natural language methods and technologies.

Content:

1. Introduction to Natural Language Processing (NLP). Basic concepts. Overview of modern applications. Linguistic resources for NLP. Corpus taxonomy. Linguistic elements of speech. Morphology. Syntactic analysis. Semantic analysis. Pragmatic.
2. Word vector representation. Simple and advanced word vector representations: word2vec, GloVe.
3. Word window classification and neural networks. Neural networks, backpropagation algorithm, neural network overfitting, regularization and activation functions. Named entity recognition. Model of neural network for the classification of grammatical phrases. Tensorflow framework.
4. Clustering for text similarity. Text clustering, similarity and distance metrics, k-means clustering, hierarchical clustering, modeling documents topics, latent semantic analysis, non-negative matrix factorization.
5. Linguistic structure. Dependency parsing. Dependency syntax and structure. Semantic analysis. A neural graph-based dependency parser.
6. Recurrent Neural Networks for language modelling. Traditional language models. Recurrent Neural Networks. Bidirectional Recurrent Neural Networks.
7. Recurrent Neural Networks based language model, their training problems and solutions. Opinion mining with Recurrent Neural Networks.
8. Recurrent Neural Networks for machine translation. Statistical machine translation methods. Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM) methods for machine translation.
9. Recursive Neural Networks. Recursive Neural Networks for structure prediction – syntax, parsing and semantic.
10. Advanced Recursive Neural Networks. Recursive Neural Networks (RNNs): Standard RNNs, Matrix-Vector RNNs, Recursive Neural Tensor Networks, Tree LSTMs and their application for paraphrase detection, relation classification, sentiment analysis, phrase similarity detection.

Assignments:

1. Word vector representation.
2. Application of classical algorithms. Word window classification.
3. Dependency parsing.
4. Development of specialized algorithms. Individual task taking into account the research topic of doctoral student. Recursive Neural Networks or Recurrent Neural Networks should be used for this task.

The course will consist of 10 lectures, 10 workshops, two home works, and two student projects.

Main literature

Steven Bird, Ewan Klein, and Edward Loper Natural Language Processing with Python – Analyzing Text with the Natural Language Toolkit, 2015, http://www.nltk.org/book

Daniel Jurafsky and James Martin (2017). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (Third edition), Prentice-Hall, Inc., 1024 p. https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf

Alexander Clark, Chris Fox and Shalom Lappin (2013). The Handbook of Computational Linguistics and Natural Language Processing, John Wiley & Sons, 802 p.
Mohamed Zakaria Kurdi (2016). Natural Language Processing and Computational Linguistics: Speech, Morphology and Syntax. Wiley-ISTE, 296 p.
CS224d: Deep Learning for Natural Language Procesing, http://cs224d.stanford.edu/syllabus.html
Benjamin Bengfort, Rebecca Bilbro, Tony Ojeda (2018) „Appplied Text Analysis with Python: Enabling language-aware data products with machine learning“
CS224n: Natural Language Processing with Deep Learning, https://web.stanford.edu/class/cs224n/

Lecturer(s) (name, surname)	Science degree	Main publications
Virginijus Marcinkevičius	Dr.	http://www.elaba.mb.vu.lt/dmsti/?aut=Virginijus+Marcinekvičius
Gražina Korvel	Dr.	http://www.elaba.mb.vu.lt/dmsti/?aut=Gražina+Korvel
Gintautas Tamulevičius	Dr.	http://www.elaba.mb.vu.lt/dmsti/?aut=Gintautas+Tamulevičius