**Introduction to Natural Language Processing**

Lecturers:

*Ekaterina Chernyak* – Lecturer, School of Data Analysis and Artificial Intelligence; Junior Research Fellow, International Laboratory of Decision Choice and Analysis

*Dmitry Ilvovsky* – Lecturer, School of Data Analysis and Artificial Intelligence, Junior Research Fellow, Laboratory for Intelligent Systems and Structural Analysis

# Course Objective

The course will introduce the basic concepts of Natural Language Processing Artificial Intelligence: part of speech tagging, latent topics, syntactic trees, text corpora and lexical databases. Methods for key word and key phrase extraction, latent topic detection, text visualization and semantic relation detection will be presented. The theoretical part of the course will be combined with the practical one: NLTK and gensim libraries and R tools for visualization will be considered.

# 2.The position of the course in the structure of the educational program

Course duration: approx. 2 weeks: 16 hours (lectures + practical studies)

Academic control forms are home assignments and an oral test test. The final mark is calculated as follows: 0.5 \* home assignments + 0.5 oral test. There will be 4 home assignments.

## 2.1.Prerequisites of the course:

* Basic linear algebra
* Basic probability theory and mathematical statistics

# 3.Topic-Wise Curricula Plan

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| № | Topic name | Course hours, total | Audience hours | |
| Lectures | Practical studies |
|  | Introduction: aims and perspective of natural language processing |  | 1 | 0 |
|  | Morphological analysis |  | 2 | 1 |
|  | Key words and key phrases extraction |  | 1 | 1 |
|  | Latent topic detection |  | 1 | 1 |
|  | Syntactical analysis |  | 1 | 1 |
|  | Text visualisation |  | 1 | 1 |
|  | Language resources |  | 1 | 1 |
|  | Distributional semantics |  | 1 | 1 |
|  | **Total** | 16 | 9 | 7 |

**4.Readings:**

1. Manning, Christopher D. *Foundations of statistical natural language processing*. Ed. Hinrich Schütze. MIT press, 1999.
2. Jurafsky, Daniel, and H. James. "Speech and language processing an introduction to natural language processing, computational linguistics, and speech." (2000).
3. Bird, Steven. "NLTK: the natural language toolkit." *Proceedings of the COLING/ACL on Interactive presentation sessions*. Association for Computational Linguistics, 2006.
4. Meyer, David, Kurt Hornik, and Ingo Feinerer. "Text mining infrastructure in R."*Journal of Statistical Software* 25.5 (2008): 1-54.
5. Řehůřek, Radim, and Petr Sojka. "Software framework for topic modelling with large corpora." (2010).