Subject code	Credits
INF5010	6

Course title in Lithuanian

KALBOS APDOROJIMO TECHNOLOGIJOS

Course title in English

NATURAL LANGUAGE TECHNOLOGIES

Short course annotation in Lithuanian (up to 500 characters)

Kalbos apdorojimo technologijos – lingvistikos ir dirbtinio intelekto sritis jungianti disciplina, kurios tikslas: išmokyti kompiuterius „suprasti“ žmonių kalbą. Semestro metu nagrinėjami populiariausi kompiuterinės lingvistikos uždaviniai; pristatomos klasikinės ir pažangiausios metodikos, leidžiančios tuos uždavinius efektyviai išspręsti ne tik anglų, bet ir norminei/nenorminei lietuvių kalbai.

Short course annotation in English (up to 500 characters)

Natural language processing is the discipline connecting linguistics and artificial intelligence, having the purpose to train computers to “understand” human language. During a semester the most popular computational linguistic tasks are analysed; classical and state-of-the-art techniques capable of solving those tasks effectively not only for English, but for the normative/non-normative Lithuanian language are presented.

Prerequisites for entering the course

Machine learning course

Course aim

This course aims to introduce fundamental techniques applied in natural language processing and to develop perception about their applicability and limitations.

Content

No.	Content (topics)
1.	Regular expressions
2.	Text pre-processing
3.	Language modelling
4.	Error correction
5.	Text classification
6.	Sentiment analysis
7.	Authorship identification
8.	Data mining
9.	Named entity recognition
10.	Syntax analysis
11.	Information retrieval
12.	Semantic analysis based on thesauruses, ontologies
13.	Question-answering systems
14.	Machine translation

Distribution of workload for students (contact and independent work hours)

Lectures	**45 hours**
Laboratory work	**15 hours**
Individual students work	**100 hours**
Total:	**160 hours**

Structure of cumulative score and value of its constituent parts

Mid-term written exam (17%), assessments of homework (33%), and final written exam (50%).

Recommended reference materials

No.	Publication year	Authors of publication and title	Publishing house	Number of copies in		
				University library	*Self-study rooms*	*Other libraries*
Basic materials						
1.	2009	Christopher D. Manning, Prabhakar Raghavan and	Cambridge University Press			http://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf

		Hinrich Schütze. *Introduction to Information Retrieval.*				
2.	2012	Dan Jurafsky, Christopher D. Manning *Natural Language Processing*	Stanford University			https://class.coursera.org/nlp/lecture

Course programme designed by

Assoc.prof. Jurgita Kapočiūtė-Dzikienė