

Dalyko kodas	Kreditai
INF5010	6

Dalyko pavadinimas lietuvių kalba

KALBOS APDOROJIMO TECHNOLOGIJOS

Dalyko pavadinimas anglų kalba

NATURAL LANGUAGE TECHNOLOGIES

Trumpa dalyko anotacija lietuvių kalba (iki 500 simbolių)

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.

Dalyko anotacija anglų kalba (iki 500 simbolių)

Natural language technologies 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 analyzed; 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.

Būtinasis pasirengimas dalyko studijoms

Mašininio mokymo kursas

Dalyko tikslas

Kurso tikslas: pristatyti pagrindines kalbos apdorojimo technologijose naudojamas metodikas ir suformuoti suvokimą apie jų taikymo galimybes bei ribotumus.

Dalyko turinys

Nr.	Turinys (temos)
1.	Reguliariosios išraiškos
2.	Pirminis teksto apdorojimas
3.	Kalbos modeliavimas
4.	Rašybos klaidų taisymas
5.	Teksto klasifikavimas
6.	Sentimentų analizė
7.	Autorystės nustatymas
8.	Informacijos gavyba
9.	Įvardintų esybių atpažinimas
10.	Sintaksinė analizė
11.	Informacijos paieška
12.	Semantinė analizė paremta tezaurais, ontologijomis
13.	Klausimų-atsakymų sistemos
14.	Mašininis vertimas

Studentų darbo krūvio paskirstymas valandomis (kontaktinio ir savarankiško darbo val.)

Paskaitos (P)	45 val.
Laboratoriniai darbai (L)	15 val.
Savarankiškas darbas	100 val.
Iš viso	160 val.

Kaupiamojo balo sandara ir jo dedamųjų svoris

Koliokvumas – 17%, namų darbas – 33%, egzaminas – 50%

Rekomenduojama literatūra

Nr.	Leidimo metai	Leidinio autoriai ir pavadinimas	Leidykla	Egzempliorių skaičius		
				<i>Universiteto bibliotekoje</i>	<i>Metodiniuose kabinetuose</i>	<i>Kitose bibliotekose</i>
Pagrindinė literatūra						
1.	2009	Christopher D. Manning, Prabhakar Raghavan and Hinrich Schütze. <i>Introduction to Information</i>	Cambridge University Press			http://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf

		<i>Retrieval.</i>				
2.	2012	Dan Jurafsky, Christopher D. Manning <i>Natural Language Processing</i>	Stanford University			https://class.coursera.org/nlp/lecture

Dalyko programos rengėjas/-ai

Doc. dr. Jurgita Kapočiūtė-Dzikienė