

DOKTORANTŪROS STUDIJŲ DALYKO SANDAS

Dalyko pavadinimas	Mokslo kryptis, kodas	Fakultetas	Katedra
Natūralios kalbos apdorojimas	Informatikos inžinerija (T 007)	Matematikos ir informatikos fakultetas	Duomenų mokslo ir skaitmeninių technologijų institutas

Studijų būdas	Kreditų skaičius ECTS	Studijų būdas	Kreditų skaičius
paskaitos	1 (pavasario semestras)	konsultacijos	1
individualus	4	seminarai	1

<p>Dalyko anotacija</p> <p>Pageidautina, kad doktorantas pasirenkantis šį dalyką išmanytų, tiesinę algebrą, matematinę analizę ir Python programavimo kalbą.</p> <p>Pagrindinis sando tikslas – gilinti natūralios kalbos apdorojimo žinias naudojant pažangius natūralios kalbos metodus ir technologijas.</p> <p>Dalyko temos</p> <ol style="list-style-type: none"> Įvadas į natūralios kalbos apdorojimą. Pagrindinės sąvokos. Šiuolaikinių naudojamų sričių apžvalga. Lingvistiniai natūralios kalbos apdorojimo ištekčiai. Garsyno taksonomija. Lingvistiniai kalbos elementai. Morfologija. Sintaksės analizė. Semantinė interpretacija. Kalbos pragmatika. Žodžių atvaizdavimas į vektorius. Paprastieji ir pažangūs atvaizdavimo metodai (word2vec, GloVe). Žodžių lango klasifikavimas ir neuroniniai tinklai. Neuroniniai tinklai, gradiento sklaidimo atgal algoritmas, neuroninio tinklo persimokymas, reguliavimas, aktyvacijos funkcijos. naudojimas atpažįstant įvardintas esybes (angl. named entity recognition). Neuroninis tinklo modelis gramatinių frazių klasifikavimui. Tensorflow programinė įranga. Kasterizavimo naudojimas natūralios kalbos apdorojime. Teksto klasifikavimas, panašumo ir atstumo metrikos, K-means kasterizavimas, Hierarchinis klasterizavimas, dokumentų grupavimas pagal temas, Latentinė semantinė analizė ir neneigiamų matricių faktorizacija. Ligvstinės struktūros. Priklausomybių analizė, priklausomybių kalba ir priklausomybių struktūra. Semantinė analizė. Neuroniniais tinklais grįsti priklausomybių analizės metodai. Rekurentiniai neuroniniai tinklai kalbos modeliavimui. Tradiciniai kalbos modeliai. Rekurentiniai neuroniniai tinklai. Dvikrypčiai rekurentiniai tinklai. Pažangūs rekurentiniai neuroniniai tinklai (GRU (angl. Gated Recurrent Unit) ir LSTM (angl. Long Short-Term Memory)) jų mokymo problemos ir sprendimai. Nuomonių tyryba su rekurentiniais tinklais. Rekurentiniai neuroniniai tinklai mašininiam vertimui. Statistiniai mašininio vertimo metodai. Neuroninių tinklų taikymas mašininiam vertimui. Rekursiniai neuroniniai tinklai. Rekursiniai neuroniniai tinklai struktūros prognozavimui – sintaksinei, gramatiniam ir semantiniam nagrinėjimui. Pažangūs rekursiniai tinklai. Rekursiniai neuroniniai tinklai (RNT): standartiniai, Matrica-vektorius RNT, Rekursiniai neuroniniai tensorių tinklai, LTSM medžiai ir jų naudojimas atpasakojimo aptikimui, sąryšių klasifikavimui, sentimentų analizei, frazių panašumo nustatymui. <p>Praktinės užduotys:</p> <ol style="list-style-type: none"> Žodžių atvaizdavimas į vektorius. Klasikinių algoritmų taikymas teksto klasifikavimui. Klasifikavimas naudojant žodžių langą. Priklausomybių nustatymas naudojant neuroninius tinklus. Specializuotų algoritmų kūrimas. Pagal doktoranto vykdomą mokslinę temą parinkta užduotis, kurios sprendimui būtų naudojami rekursiniai ar rekurentiniai neuroniniai tinklai. <p>Sandą sudarys, 10 paskaitų, 10 seminarų, du atsiskaitomieji darbai (projektai), du namų darbai..</p> <p>Pagrindinė literatūra</p> <p>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</p>

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 Processing, http://cs224d.stanford.edu/syllabus.html
Benjamin Bengfort, Rebecca Bilbro, Tony Ojeda (2018) „Applied 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/

Konsultuojančiųjų dėstytojų vardas, pavardė	Mokslo laipsnis	Svarbiausieji darbai mokslo kryptyje, paskelbti per pastaruosius 5 metus
Virginijus Marcinkevičius	Dr.	http://www.elaba.mb.vu.lt/dmsti/?aut=Virginijus+Marcinkevičius
Grażina Korvel	Dr.	http://www.elaba.mb.vu.lt/dmsti/?aut=Grażina+Korvel
Gintautas Tamulevičius	Dr.	http://www.elaba.mb.vu.lt/dmsti/?aut=Gintautas+Tamulevičius