



<b>Назив предмета:</b> ОБРАДА ПРИРОДНИХ ЈЕЗИКА		
<b>Статус предмета:</b> изборни		
<b>Број ЕСПБ:</b> 10		
<b>Услов:</b> Уписан одговарајући семестар		
<b>Циљ предмета</b> Овај предмет треба да омогући студентима да стекну дубља и потпунија знања о техникама и алгоритмима који се примењују у обради и анализи текста и могућностима њихове примене у практичним ситуацијама. Предмет треба да представља припрему студената за самосталан истраживачки рад у области обраде природног текста коришћењем стандардних статистичких приступа, као и приступа који се заснивају на дубоком учењу.		
<b>Исход предмета</b> Студенти познају различите напредне методе и технике које се употребљавају за обраду и анализу текста. Студенти имају способност да коришћењем научне методологије врши анализу, синтезу и предвиђање решења и њихових последица на циљне проблеме. Студенти су оспособљени да напредне технике дубоког учења примене у решавању комплексних реалних проблема који се тичу обраде текста са различитих аспеката. Студенти су компетентни да врше даља истраживања у овој области.		
<b>Садржај предмета</b> Сегментација текста; препознавање речи и препознавање реченица. Корпуси и токенизација. Морфолошка анализа текста – лематизација, етикетирање врсте речи, избор скупа етикета, стеминг. Статистички модели језика. Нормализација текста. Отклањање вишезначности. Синтаксна анализа и парсирање. Векторска репрезентација речи, врећа речи, TF-IDF. Лексичка семантика. Детекција ентитета, анализа сентимента, екстракција термина, двосмисленост, WordNet. Релациона семантика. Екстракција везе између ентитета, семантичко парсирање, означавање семантике у реченици. Обрада текста помоћу дубоких рекурентних неуронских мрежа, LSTM ћелија. Трансформерс модели, BERT, ViLBERT, RoBERT, ALBERT, Distil BERT итд. Примене: сумаризација текста, машинско превођење, генерисање природног језика, разумевање природног језика, питања и одговори, чет-ботови.		
<b>Препоручена литература</b> 1. Chris Manning and Hinrich Schütze. <i>Foundations of Statistical Natural Language Processing</i> . MIT Press. Cambridge, MA: 1999. 2. Ian Goodfellow, Yoshua Bengio and Aaron Courville. <i>Deep learning</i> . MIT Press, 2016. 3. Daniel Jurafsky & James H. Martin. <i>Speech and Language Processing</i> . Prentice Hall 2008. 4. Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta, Harshit Surana. <i>Practical Natural Language Processing</i> . O'Reilly Media, 2020. 5. Ruslan Mitkov. <i>The Oxford Handbook of Computational Linguistics</i> , 2 <sup>nd</sup> edition. Oxford Univ. Press, 2014. 6. B. Liu. <i>Sentiment Analysis: Mining Opinions, Sentiments, and Emotions</i> . Cambridge University Press, 2015.		
Број часова активне наставе:	Теоријска настава: 5	Студијски истраживачки рад: 2
<b>Методe извођења наставe</b> Предавања подржана практичним примерима из обраде и анализе природног језика. Самостални рад студената уз одговарајућу литературу, консултације.		
<b>Оцена знања (максимални број поена 100)</b> Семинарски рад: 40 поена, Писмени испит: 30 поена, Усмени испит: 30 поена		