

Студијски програм: Основне академске студије информатика				
Назив предмета: Обрада природног језика и говора				
Наставник: Јелена Граовац, Ивана Танасијевић				
Статус предмета: изборни				
Број ЕСПБ: 6				
Услов: Машинско учење, Дубоко учење				
Циљ предмета: Преглед метода у обради природног језика и говора заснованих на моделирању знања о језику и на статистичким моделима природних језика. Увођење студента у проблеме морфолошке, синтаксичке и семантичке обраде природних језика и методе њиховог решавања.				
Исход предмета: Студент познаје проблеме анализе природног језика и говора и методе њиховог решавања, са акцентом на обради српског језика.				
Садржај предмета: <ul style="list-style-type: none"> - Сегментација текста - препознавање речи и реченица. Нормализација текста, n-грамска репрезентација текста. - Граматике. Речници. Лексикони. Wordnet. - Корпуси. Паралелни и упоредиви корпуси. Вишејезичност. - Пробабилистички језички модели. - Класификација текста и препознавање сентимента. - Векторска семантика (енг. word embeddings). - Етикетање врсте речи и именованих ентитета. - Неуронски модели за обраду текста и говора. - Трансформери и преобучени велики језички модели. Фино подешавање модела (енг. fine-tuning). - Машинско превођење. - Упитнички системи и екстракција информација. - Chatbot-ови и системи за симулацију дијалога. - Аутоматско препознавање говора и претварање текста у говор. 				
Литература: <ol style="list-style-type: none"> 1. J. Graovac. Obrada prirodnog jezika i govora. Skripta u pripremi. 2. Dan Jurafsky and James H. Martin. Speech and Language Processing. Stanford Univ. 2023. URL: https://web.stanford.edu/~jurafsky/slp3/ed3book_jan72023.pdf 3. Ruslan Mitkov, The Oxford Handbook of Computational Linguistics, Oxford University Press, 2003. 4. Daniel Jurafsky & James H. Martin, Speech and Language Processing, Prentice Hall, New Jersey, 2000. 5. S. Nirenburg et al: Ontological Semantics, MIT Press, 2004. 4. Kenneth R. Beesley and Lauri Karttunen <p>Наставник може изабрати другу одговарајућу актуелну литературу.</p>				
Бр. час. акт. наставе: 5	Теоријска настава: 3	Прак. настава: 2	Лаб.вежбе: -	СИР: -
Методе извођења наставе: Фронтални, групни, индивидуални и практични.				
Оцена знања (максималан број поена је 100)				
Предиспитне обавезе	поена	Завршни испит	поена	
активност у току предавања	20	писмени испит	-	
практична настава	-	усмени испит	-	
колоквијум-и	-	писмено-усмени испит	50	
семинар-и	30			