

NAZIV PREDMETA		Uvod u obradu prirodnog jezika				
Kod	PMII40	Godina studija	DS-1			
Nositelj/i predmeta	doc.dr. sc. Branko Žitko	Bodovna vrijednost (ECTS)	5,0			
Suradnici		Način izvođenja nastave (broj sati u semestru)	P	S	V	T
			30		30	
Status predmeta	obavezan predmet	Postotak primjene e-učenja				
OPIS PREDMETA						
Ciljevi predmeta	<p>Upoznavanje područja obrade prirodnog jezika.</p> <p>Obrada na morfološkom, sintaktičkom, semantičkom i pragmatičkom nivou s lingvističke i računalne perspektive.</p> <p>Usvajanje temeljnih modela i algoritama za obradu prirodnog jezika.</p>					
Uvjeti za upis predmeta i ulazne kompetencije potrebne za predmet	Ulazne kompetencije: linearna algebra i vjerojatnost, objektno orijentirano programiranje u Pythonu.					
Očekivani ishodi učenja na razini predmeta (4-10 ishoda učenja)	<p>opisati metode obrade prirodnog jezika</p> <p>pripremiti podatke za obradu prirodnog jezika</p> <p>implementirati algoritme za obradu prirodnog jezika u Pythonu</p> <p>ocijeniti i usporediti rezultate obrade prirodnog jezika</p>					
Sadržaj predmeta detaljno razrađen prema satnici nastave	<p>Tjedan1:</p> <p>Predavanja: Uvodno predavanje: nastavnici, obaveze studenata, elementi tekućeg praćenja, ispit, ocjena, predstavljanje ciljeva kolegija, literatura</p> <p>Problemi obrade prirodnog jezika, težinska kategorizacija</p> <p>Vježbe: Rad s tekstualnim datotekama i Web sadržajem u programskom jeziku Python</p> <p>Tjedan2:</p> <p>Predavanja: Osnove obrade jezika, regularni izrazi, tokenizacija riječi, normalizacija riječi i izvlačenje korijena, segmentacija rečenice i stabla odluke</p> <p>Vježbe: Regularni izrazi u Pythonu, segmentacija riječi i rečenica pomoću regularnih izraza, izvlačenje teksta iz Web stranice</p> <p>Tjedan3:</p> <p>Predavanja: Minimalna udaljenost dva niza znakova, izračun minimalne udaljenosti, povratno praćenje kod izračuna poravnanja, težinska minimalna udaljenost, minimalna udaljenost u računalnoj biologiji</p> <p>Vježbe: Implementacija algoritam minimalne udaljenosti u Pythonu, Implementacija povratnog praćenja kod izračuna minimalne udaljenosti</p> <p>Tjedan4:</p> <p>Predavanja: Modeliranje jezika, n-gram, procjena vjerojatnosti n-grama, evaluacija i perpleksija, generalizacija i problem nula, Laplaceovo izgladivanje, interpolacija i odustajanje, good-turing izgladivanje</p> <p>Vježbe: Modeliranje n-grama u Pythonu, generiranje jezika temeljem n-grama.</p> <p>Tjedan5:</p> <p>Predavanja: Ispravljanje pravopisnih grešaka, kanal sa šumom, greške stvarnih riječi</p> <p>Vježbe: Implementacija algoritma za ispravljanje pravopisnih grešaka temeljem rječnika i minimalne udaljenosti</p> <p>Tjedan6:</p> <p>Predavanja: Klasifikacija teksta, naivni Bayes klasifikator, formalizacija naivnog Bayesovog klasifikatora, učenje klasifikatora, odnos s modelom jezika,</p>					

	<p>multinominalni naivni Bayesov klasifikator, preciznost i opoziv, F-mjera, evaluacija klasifikatora Vježbe: Implementacija naivnog Bayesovog klasifikatora u Pythonu Tjedan7: Predavanja: Sentimentalna analiza i osnovni algoritmi, leksikon sentimentata, učenje leksikona sentimentata Vježbe: Implementacija sentimentalnog analizatora u Pythonu zasnovanog na leksikonu Tjedan8: Predavanja: usporedba diskriminativnog i generativnog klasifikatora, osobine diskriminativnog klasifikatora, linearni klasifikator temeljem na osobinama teksta, model maksimalne entropije, optimizacija vjerojatnosti Vježbe: Tjedan9: Predavanja:Prepoznavanje imenovanih entiteta (NER), ekstrakcija informacija, evaluacija NER, slijedni model za NER, maksimalna entropija slijednog modela Vježbe: Implementacija Markovljevog modela maksimalne entropije za identifikaciju naziva osoba Tjedan10: Predavanja: Ekstrakcija relacija (RE), korištenje uzoraka kod RE, nadzirana RE, polunadzirana i nenadzirana RE Vježbe: Treniranje Markovljevog modela maksimalne entropije Tjedan11: Predavanja: Parsiranje teksta, sintaktičke strukture, empirijski pristup parsiranju, eksponencijalni problem parsiranja Vježbe: Modeliranje CFG Tjedan12: Predavanja: Probabilističko parsiranje, kontekstno neovisne gramatike (CFG) i probabilističke kontekstno neovisne gramatike (PCFG), transformacija gramatike, CKY parsiranje Vježbe: Implementacija CKY parsera za PCFG Tjedan13: Predavanja: Vraćanje informacija IR, term-dokument matrice, invertirani indeksi, obrada upita temeljem invertiranih indeksa, frazni upiti i pozicijski indeksi, rangirani IR, bodovanje, TF-IDF težine, model vektorskog prostora Vježbe: Implementacija IR sustava temeljenog na TF-IDF Tjedan14: Predavanja: Sustavi za odgovaranje na pitanja QA, tipovi odgovora i formuliranje upita, ekstrakcija odgovora, korištenje znanja u QA Vježbe: Implementacija QA sustava korištenjem Wiki resursa Tjedan15: Predavanja: Uvod u sumarizaciju, generiranje isječaka i ekstrakcija odgovora, evaluacija sumarizacije, sumarizacija skupa dokumenata Vježbe: Implementacija sumarizacije temeljena na ekstrakciji informacija</p>	
Vrste izvođenja nastave:	<input checked="" type="checkbox"/> predavanja <input type="checkbox"/> seminari i radionice <input checked="" type="checkbox"/> vježbe <input type="checkbox"/> on line u cijelosti <input type="checkbox"/> mješovito e-učenje <input type="checkbox"/> terenska nastava	<input checked="" type="checkbox"/> samostalni zadaci <input type="checkbox"/> multimedija <input type="checkbox"/> laboratorij <input type="checkbox"/> mentorski rad <input type="checkbox"/> domaće zadaće
Obveze studenata	pohađanje nastave aktivno sudjelovanje u nastavnom procesu pismeni ispit	

Praćenje rada studenata (<i>upisati udio u ECTS bodovima za svaku aktivnost tako da ukupni broj ECTS bodova odgovara bodovnoj vrijednosti predmeta</i>):	Naziv	Ects	Naziv	Ects	Naziv	Ects
	Pohađanje nastave	2	Istraživanje		Eksperimentalni rad	
	Usmeni ispit		Referat		Domaće zadaće	
	Seminarski rad		Esej			
	Kolokvij		Praktični rad	1		
	Pismeni ispit	2	Projekt			
Ocjenjivanje i vrjednovanje rada studenata tijekom nastave i na završnom ispitu	Aktivnost studenata na predavanjima i vježbama (prisutnost na vježbama, rješavanje zadataka) (50 %). Pismeni dio ispita (50 %), Završna ocjena izvodi se na temelju svih navedenih ocjena.					
Obvezna literatura (dostupna u knjižnici i putem ostalih medija)	Naslov			Broj primjeraka u knjižnici	Dostupnost putem ostalih medija	
	D. Jurafsky, J. H. Martin, (2000) Speech and Language Processing, PrenticeHall			1		
Dopunska literatura	S. Bird, E. Klain, E. Looper, (2009) Natural Language Processing with Python, O'Reilly Media					
Načini praćenja kvalitete koji osiguravaju stjecanje utvrđenih ishoda učenja	razgovor sa studentima studentska evaluacija primjenom anonimne ankete uspjeh studenata na ispitu samoprocjena.					
Ostalo (prema mišljenju predlagatelja)						