

Sveučilište u Rijeci
ODJEL ZA INFORMATIKU
Radmile Matejčić 2, Rijeka
Akademska 2018./2019. godina

RAČUNALNA ANALIZA PRIRODNOG JEZIKA

Studiji: Diplomski studij informatike IKS OBAVEZNI + PI IZBORNI
ECTS bodovi: 6
Nastavno opterećenje: 2 + 2
Web stranica predmeta: <https://moodle.srce.hr/>
<http://www.inf.uniri.hr/hr/nastava/izvedbeni-program/54-hr/nastava/izvedbeni-program/250-jednopredmetni-diplomski-studij-informatike-modul-informacijski-i-komunikacijski-sustavi.html>

Studiji: Diplomski studij informatike (dvopredmetni) OBAVEZNI
ECTS bodovi: 3
Nastavno opterećenje: 2 + 2
Web stranica predmeta: <https://moodle.srce.hr/>
<http://www.inf.uniri.hr/hr/nastava/izvedbeni-program/54-hr/nastava/izvedbeni-program/254-dvopredmetni-diplomski-studij.html>

Nositeljica predmeta:

prof. dr. sc. Sanda Martinčić-Ipšić
e-mail: smarti@inf.uniri.hr
Ured: Radmile Matejčić 2, Rijeka, 409
Vrijeme konzultacija: četvrtkom po prethodnom dogovoru e-mailom

Asistent:

Karlo Babić
e-mail: karlo.babic [at] inf.uniri.hr
Ured: Ulica Radmile Matejčić 2, 419
Vrijeme konzultacija: srijedom po dogovoru e-mailom

RAČUNALNA ANALIZA PRIRODNOG JEZIKA

Razvijanje općih i specifičnih kompetencija (znanja i vještina)

Studenti trebaju steći temeljna znanja o postupcima računalne analize prirodnog jezika i računalne lingvistike.

Korespondentnost i korelativnost programa

Program kolegija povezan je sa kolegijima Formalni jezici i jezični procesori 1.

Okvirni sadržaj predmeta

Uvod u područje i pregled pojmova računalne analiza prirodnog jezika. Računalna lingvistika i jezične tehnologije. Jezični resursi, korpusi, rječnici, leksikoni.

Prepoznavanje sintakse. Vjerojatnostni modeli. N-gramski modeli. Perpleksnost. Jezični alati za provjeru gramatike i pravopisa. Morfološki analizatori. Obilježivači vrsta riječi. Parsiranje pomoću kontekstno neovisnih gramatika.

Prepoznavanje semantike. Prepoznavanje leksičkoga i rečeničkoga značenja.

Prepoznavanje pragmatike. Diskurs. Dijalog. Generiranje jezika.

Identifikacija jezika. Strojno prevođenje i strojno potpomognuto prevođenje.

Pretraživanje dokumenata. Strukture indeksiranja dokumenata.

Oblici provođenja nastave i način provjere znanja

Predavanja, vježbe, samostalni zadaci, konzultacije, seminarski rad, E-učenje

Popis literature potrebne za studij i polaganje ispita

1. D. Jurafsky, J. H. Martin: Speech and Language Processing, An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition, Prentice Hall, 2000.
2. C. Manning, H. Schütze: Foundations of Statistical NLP, MIT Press, Cambridge, Massachusetts, 1999.

Popis literature koja se preporučuje kao dopunska

1. J. Allen: Natural Language Understanding
2. Minker, W., Bennacef, S. Speech and human-Machine dialog. Kluwer Academic Publishers. Boston. 2004.
3. P. Nugues An Introduction to Language Processing with Perl and Prolog, Springer, 2006
4. http://www.cs.lth.se/home/Pierre_Nugues/ilppp/slides.shtml
5. M. Covington, Natural Language Processing for Prolog Programmers, Prentice Hall 1994.
6. S. Bird, E. Klein, E. Loper: Natural Language Processing with Python, O'Riley 2009

Način praćenja kvalitete i uspješnosti izvedbe predmeta

Kroz ustrojeni sustav osiguranja kvalitete Odjela za informatiku.

Preduvjeti za upis predmeta: nema

Mogućnost izvođenja na stranom jeziku: DA

R. BR.	OČEKIVANI ISHODI
1.	Definirati i objasniti postupke i pojmove računalne analize prirodnog jezika.
2.	Razumjeti postupke i standarde izgradnje jezičnih resursa.
3.	Izraditi postupke za izgradnju i analizu jezičnih resursa.
4.	Objasniti postupke morfološke analize teksta te obilježavanja vrste riječi.
5.	Izgraditi statistički jezični model te definirati i razumjeti postupke izgradnje.
6.	Razumjeti postupke Early-evog i CYK parsiranja.
7.	Objasniti i definirati formalizme za zapis značenja u jezičnim aplikacijama.
8.	Razumjeti postupke semantičke analize i ekstrakcije podataka iz nestrukturiranih izvora.
9.	Objasnite postupke automatske klasifikacije teksta i naprednog pretraživanja dokumenta.
10.	Nabrojati i definirati postupke analize diskursa.
11.	Objasniti modele i strukturu sustava za automatsko vođenje dijaloga s računalom te ih primijeniti u samostalnom radu.
12.	Objasniti principe strojno potpomognutog prevođenja i/ili određivanja polariteta teksta.

AKTIVNOSTI I OCJENJIVANJE STUDENATA

VRSTA AKTIVNOSTI	ECTS	ISHODI UČENJA	SPECIFIČNA AKTIVNOST	METODA PROCJENJIVANJA	BODOVI MAX.
Prisutnost na nastavi	1	1-12	Prisutnost studenata	Popisivanje (evidencija)	0
Samostalne vježbe na računalu	0,5	3-8	Samostalni zadaci na vježbama	0-5 bodova po zadatku	15
Kontinuirana provjera znanja	1,5	1-12	Jedna kontrolna zadaća (kolokviji)	0-25 bodova za zadaću, ovisno o stupnju točnosti i potpunosti	25
Samostalni zadaci	2	2-12.	Izrada praktičnog seminarskog rada	0-30 bodova za seminarski rad, ovisno o pravilnosti realiziranog postupka ali i samog postupka izrade	30
Završni ispit	1	1-12	Pisani test	0-30 bodova, ovisno o stupnju točnosti i potpunosti	30
UKUPNO	6				100

Obveze i vrednovanje studenata

1. Pohađanje nastave
Pohađanje nastave je obavezno i vodi se evidencija pohađanja za svakoga studenta. Predavanja i vježbe se izvode u blokovima od 2 sata prema tablici u nastavku.
2. Samostalni zadaci na računalu
Tijekom semestra studenti će izrađivati samostalne zadatke vezane uz sadržaj vježbi, na kojima je moguće prikupiti do 15 bodova.
3. Seminarski rad
Tijekom semestra studenti će izrađivati praktični (projektni) seminarski rad te ga prezentirati na roku napisanom u kalendaru izvođenja. Seminarskim radom moguće sakupiti do 30 bodova.
4. Kontrolna zadaća (kolokviji)
Tijekom semestra provjerit će se znanje studenta kroz kontrolnu zadaću (kolokvij), na kojem je moguće prikupiti do 25 bodova, nema praga.
5. Dvopredmetni studenti nisu dužni izraditi samostalne zadatke, ali s predajom mogu prikupiti dodatnih max 20 bodova, koji se normiraju nakon izvršenih obveza na kolegiju, pri čemu uvjet za pristupanje završnom i popravnom ispitu ostaje nepromijenjen.
6. **Nepoštivanje postavljenih rokova smatra se neizvršavanjem obveza.**

Završni ispit

Kontinuiranim radom tijekom semestra na prethodno opisani način studenti mogu ostvariti najviše 70 ocjenskih bodova, a da bi mogli pristupiti završnom ispitu moraju ostvariti 50% i više bodova (minimalno 35).

Studenti koji su skupili najmanje 35 ocjenskih bodova, mogu pristupiti završnom ispitu.

Završni ispit nosi udio od maksimalno 30 ocjenskih bodova, a smatra se položenim samo ako na njemu student postigne minimalno 50%-ni uspjeh (ispitni prag je 50% uspješno riješenih zadataka).

Ukoliko je završni ispit prolazan, skupljeni bodovi će se pribrojati prethodnima i prema ukupnom rezultatu formirati će se pripadajuća ocjena. U suprotnom, student ima pravo pristupa završnom ispitu još 2 puta (ukupno do 3 puta).

Konačna ocjena

Računalna analiza prirodnog jezika

Donosi se na osnovu zbroja svih bodova prikupljenih tijekom izvođenja nastave prema sljedećoj skali:

A – 90% - 100%	(ekvivalent: izvrstan 5)
B – 75% - 89,9%	(ekvivalent: vrlo dobar 4)
C – 60% - 74,9%	(ekvivalent: dobar 3)
D – 50% - 59,9%	(ekvivalent: dovoljan 2)
F – 0% - 49,9%	(ekvivalent: nedovoljan 1)

RASPORED NASTAVE U AKADEMSKOJ GODINI 2018./2019.

Nastava će se odvijati u ljetnom semestru prema sljedećem rasporedu:

Predavanja	četrvtak	12-14	357
Vježbe	četrvtak	10-12	365

Izvođenje nastave prikazano je u sljedećoj tablici:

KALENDAR PREDAVANJA I VJEŽBI – ljetni (IV) semestar ak. godine 2018./2019.

Datum	Tema	Nastava	Izvođač
7.3.	Uvod, opće informacije, pregled pojmova	P	Dr.sc. Sanda Martinčić-Ipšić
7.3.	Jezični resursi, korpusi, rječnici, leksikoni.	P	Dr.sc. Sanda Martinčić-Ipšić
14.3.	Word-net, korpusi, alati za korpus	P	Dr.sc. Sanda Martinčić-Ipšić
14.3.	Web scraping	V	Karlo Babić
21.3.	Morfološki analizatori. Obilježivači vrsta riječi. Lematizatori	P	Dr.sc. Sanda Martinčić-Ipšić
21.3.	NLTK 1 - Uvod	V	Karlo Babić
28.3.	Statistički jezični modeli. N-gramski modeli. Perpleksnost	P	Dr.sc. Sanda Martinčić-Ipšić
28.3.	NLTK 2 – Rad s korpusima	V	Karlo Babić
4.4.	Parsiranje. Prepoznavanje značenja.	P	Dr.sc. Sanda Martinčić-Ipšić
4.4.	NLTK 5 – Automatsko dodjeljivanje POS oznaka	V	Karlo Babić
11.4.	Pretraživanje dokumenata. Strukture indeksiranja dokumenata	P	Dr.sc. Sanda Martinčić-Ipšić
11.4.	NLTK 4 – Stvaranje kategoriziranih korpusa	V	Karlo Babić
18.4.	Vektorski prostor. Klasifikacija dokumenata.	P	Dr.sc. Sanda Martinčić-Ipšić
18.4.	NLTK 3 – Klasifikacija tekstova	V	Karlo Babić
	rok za prijavu teme za seminar 20.4.		
	Kolokvij 25.4.		
2.5.	Otkrivanje stavova i mišljenja iz teksta	P	Dr.sc. Sanda Martinčić-Ipšić
2.5.	NLTK 6 – Pretraživanje i dohvaćanje informacija	V	Karlo Babić
9.5.	Text Mining - Dubinska analiza teksta	P	Dr.sc. Sanda Martinčić-Ipšić
9.5.	Text mining 1		
16.5.	Strojno učenje, klasifikatori	P	Dr.sc. Sanda Martinčić-Ipšić
16.5.	Text mining 2	V	Karlo Babić
23.5.	Dubinsko strojno učenje za tekst	P	Dr.sc. Sanda Martinčić-Ipšić
23.5.	Deep/Text mining 3	V	Karlo Babić
30.5.	Industrijska Radionica	V	Karlo Babić
30.5.	Industrijska Radionica	V	Karlo Babić
6.6.	Diskurs. Dijalog.	P	Dr.sc. Sanda Martinčić-Ipšić
6.6.	Deep/Text mining 4	V	Karlo Babić
	Rok za predaju seminara 7.6.		
13.6.	Seminari	P	Dr.sc. Sanda Martinčić-Ipšić
13.6.	Seminari	V	Karlo Babić

P – predavanja V – vježbe

Ispitni rokovi:

27.06.2019. 10:00 četvrtak
 11.07.2019. 10:00 četvrtak
 5.09.2019 10:00 - četvrtak
 12.09.2019. 10:00 - četvrtak