

Sveučilište u Rijeci  
ODJEL ZA INFORMATIKU  
Radmile Matejčić 2, Rijeka  
Akademska 2014./2015. godina

## RAČUNALNA ANALIZA PRIRODNOG JEZIKA

Studiji: Diplomski studij informatike IKS OBAVEZNI + PI IZBORNI  
ECTS bodovi: 5  
Nastavno opterećenje: 2 + 2  
Web stranica predmeta: <http://mudri.uniri.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: <http://mudri.uniri.hr>  
<http://www.inf.uniri.hr/hr/nastava/izvedbeni-program/54-hr/nastava/izvedbeni-program/254-dvopredmetni-diplomski-studij.html>

### Nositeljica predmeta:

izv. prof. dr. sc. Sanda Martinčić-Ipšić  
e-mail: [smarti@inf.uniri.hr](mailto:smarti@inf.uniri.hr)  
Ured: Radmile Matejčić 2, Rijeka, 409  
Vrijeme konzultacija: utorkom po prethodnom  
dogovoru e-mailom

### Asistent:

Luka Vretenar  
e-mail: [lvretenar@inf.uniri.hr](mailto:lvretenar@inf.uniri.hr)  
Ured: Radmile Matejčić 2, Rijeka, 419  
Vrijeme konzultacija: srijedom po prethodnom  
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](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	1-12	Jedna kontrolna zadaća (kolokviji)	0-25 bodova za zadaću, ovisno o stupnju točnosti i potpunosti	25
Samostalni zadaci	1,5	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
<b>UKUPNO</b>	<b>5</b>				<b>100</b>

### 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.

#### 5. Završni ispit

Na završnom ispitu studenti moraju skupiti barem 50% bodova za prolaz.

#### 6. Nepoštivanje postavljenih rokova smatra se neizvršavanjem obveza.

### OCJENJIVANJE

Studenti ukupno mogu skupiti 100 ocjenskih bodova.

- Studenti koji su skupili **manje od 40** ocjenskih bodova moraju ponovno upisati kolegij.
- Studenti koji su skupili **od 40-49** ocjenskih bodova pristupaju popravnom ispitu na kojem moraju ostvariti minimalno 50% bodova. Studenti koji zadovolje na popravnom ispitu dobivaju ocjenu E bez obzira na stupanj postignuća tj. broj bodova. Studenti koji ne zadovolje moraju ponovo upisati predmet.
- Studentima koji su skupili najmanje 50 ocjenskih bodova kontinuiranim radom na nastavnim aktivnostima završna ocjena se donosi se na osnovu zbroja svih bodova prema sljedećoj skali:

A – 90% - 100% (ekvivalent: izvrstan 5)

B – 80% - 89% (ekvivalent: vrlo dobar 4)

C – 70% - 79% (ekvivalent: dobar 3)

D – 60% - 69% (ekvivalent: dovoljan 2)

E – 50% - 59% (ekvivalent: dovoljan 2)

**RASPORED NASTAVE U AKADEMSKOJ GODINI 2014./2015.**

Nastava će se odvijati u ljetnom semestru prema sljedećem rasporedu:

<b>Predavanja</b>	<b>utorkom</b>	<b>10-12</b>	<b>028</b>
<b>Vježbe</b>	<b>srijedom</b>	<b>12-14</b>	<b>366</b>

Izvođenje nastave prikazano je u sljedećoj tablici:

**KALENDAR PREDAVANJA I VJEŽBI – ljetni (IV) semestar ak. godine 2014/2015.**

Datum	Tema	Nastava	Izvođač
3.3.2015.	Uvod , opće informacije, pregled pojmova	P	Dr.sc. Sanda Martinčić-Ipšić
4.3.2015.	Jezični resursi, korpusi, rječnici, leksikoni.	P	Dr.sc. Sanda Martinčić-Ipšić
10.3.2015.	<a href="#">Word-net, korpusi, alati za korpus</a>	V	Luka Vretenar
11.3.2015.	<a href="#">Python, IPython</a>	V	Luka Vretenar
17.3.2015.	Morfološki analizatori. Obilježivači vrsta riječi. Lematizatori	P	Dr.sc. Sanda Martinčić-Ipšić
18.3.2015.	<a href="#">Python, regularni izrazi</a>	V	Luka Vretenar
24.3.2015.	Statistički jezični modeli. N-gramski modeli. Perpleksnost.	V	Luka Vretenar
25.3.2015.	<a href="#">Web scraping</a>	V	Luka Vretenar
31.3.2015.	Parsiranje. Prepoznavanje značenja.	P	Dr.sc. Sanda Martinčić-Ipšić
1.4.2015.	<a href="#">NLTK 1</a>	V	Luka Vretenar
7.4.2015.	Pretraživanje dokumenata. Strukture indeksiranja dokumenata	P	Dr.sc. Sanda Martinčić-Ipšić
8.4.2015.	<a href="#">NLTK 2</a>	V	Luka Vretenar
14.4.2015.	Vektorski prostor. Latentno indeksiranje. Klasifikacija dokumenata	P	Dr.sc. Sanda Martinčić-Ipšić
15.4.2015.	<a href="#">NLTK 3</a>	V	Luka Vretenar
21.4.2015.	Otkrivanje stavova i mišljenja iz teksta	P	Dr.sc. Sanda Martinčić-Ipšić
24.4.2015.	<a href="#">NLTK 4</a>	V	Luka Vretenar
28.4.2015.	<b>Kolokvij</b>	P	Dr.sc. Sanda Martinčić-Ipšić
29.4.2015.	<a href="#">NLTK 5</a>	V	Luka Vretenar
5.5.2015.	Dubinska analiza teksta.	P	Dr.sc. Sanda Martinčić-Ipšić
6.5.2015.	<a href="#">Machine learning, klasifikatori</a>	V	Luka Vretenar
12.5.2015.	<a href="#">Machine learning, klasifikatori</a>	V	Luka Vretenar
13.5.2015.	Industrijska Radionica	V	Luka Vretenar
19.5.2015.	Sustavi preporuka.	P	Dr.sc. Sanda Martinčić-Ipšić
20.5.2015.	<a href="#">Orange- text mining 1</a>	V	Luka Vretenar
26.5.2015.	Page Rank HITS	P	Dr.sc. Sanda Martinčić-Ipšić
27.5.2015.	<a href="#">Orange- text mining 2</a>	V	Luka Vretenar
2.6.2015.	Prepoznavanje pragmatike. Diskurs. Dijalog.	P	Dr.sc. Sanda Martinčić-Ipšić
3.6.2015.	Seminari	V	Luka Vretenar
9.6.2015.	Seminari	P	Dr.sc. Sanda Martinčić-Ipšić
10.6.2015.	Seminari	V	Luka Vretenar

P – predavanja V – vježbe

**Ispitni rokovi:**

18.06.2015. 10:00  
 02.07.2015. 10:00  
 03.09.2015. 10:00  
 10.09.2015. 10:00