

Podiplomski magistrski študijski program 2. stopnje PODATKOVNA ZNANOST

Opisi predmetov

OBVEZNI PREDMETI ŠTUDIJSKEGA PROGRAMA

Ime predmeta: **ETIKA V PODATKOVNI ZNANOSTI**

Število ECTS kreditnih točk: **3**

Vsebina:

Sodobne tematika etičnih pristopov k upravljanju in raziskovanju podatkov. Možne vsebine vključujejo:

- Osnovni temelji etike
- Informirano soglasje
- Lastništvo podatkov
- Zasebnost
- Anonimnost
- Veljavnost podatkov
- 'Poštenost' algoritmov
- Družbene posledice
- Etični kodeks

Ime predmeta: **INTELIGENTNI SISTEMI**

Število ECTS kreditnih točk: **6**

Vsebina:

Pri predmetu bodo obravnavana poglavja, ki lahko med drugim vključujejo (a niso omejena le na) naslednje teme glede na potrebe in razvoj širšega področja predmeta:

Glavne tematike – poglavja (s podpoglavji):

- Temeljne vsebine [jedro*]
- Osnovne strategije preiskovanja [jedro]
- Sklepanje na podlagi znanja [jedro]
- Napredne strategije preiskovanja [opcija*]
- Napredno sklepanje [opcija]
- Agenti [opcija]
- Obdelava naravnih jezikov [opcija]
- Strojno učenje [opcija]
- Sistemi za načrtovanje [opcija]
- Robotika [opcija]
- Zaznavanje [opcija]

* Jedro predmeta tvorijo prva 3 poglavja; nadaljnjih 8 je opcijskih. Izmed opcijskih bodo vsako leto v okviru tega predmeta obdelana 3 - raziskovalno najaktualnejša - poglavja.

Ime predmeta: **IZBRANA POGLAVJA IZ DISKRETNE MATEMATIKE**

Število ECTS kreditnih točk: **6**

Vsebina:

Predavajo se najpomembnejše raziskovalno aktualne teme iz področja diskretne matematike, ki med drugimi lahko vključujejo naslednja vsebinska podpodročja:

- Teorija konfiguracij
- Teorija grafov
- Algebraične metode v teoriji grafov,
- Teorija velikih omrežij in analiza,
- Učenje na omrežjih,
- Slučajni sprehodi na grafih,
- Svetovni splet kot graf.

Pri SE se bodo obravnavali konkretni primeri problemov, s katerimi se srečujemo pri izvajanju aplikativnih projektov v okviru raziskovalnega programa P1-0285 (npr. analiza podatkov v okviru EEG meritev).

Ime predmeta: **IZBRANA POGLAVJA IZ VIZUALIZACIJE PODATKOV**

Število ECTS kreditnih točk: **6**

Vsebina:

Pri predmetu bodo obravnavana poglavja, ki lahko med drugim vključujejo (a niso omejena le na) naslednje teme glede na potrebe in razvoj širšega področja predmeta:

- ustrezna razporeditev vizualnih atributov do spremenljivk podatkov,
- oblikovanje z barvo in kontrastom in osvetljenostjo,
- psihologija človeškega vida in zaznavanja,
- vizualne analitike,
- interakcija,
- pripovedovanje zgodb,
- besedilna vizualna analitika,
- vizualizacija masovnih podatkov,
- vizualizacija negotovosti,
- vizualizacija omrežij in mrež,
- kartografski prikaz,
- animacija in časovna vizualizacija.

Ime predmeta: **PODATKOVNE BAZE ZA MASOVNE PODATKE**

Število ECTS kreditnih točk: **6**

Vsebina:

- Relacijske podatkovne baze
- podatkovne baze NoSQL: dokumentne podatkovne baze; grafske podatkovne baze; stolpične podatkovne baze; porazdeljene podatkovne baze.
- Optimizacija poizvedb.
- Kontrola sočasnega izvajanja.
- Varovanje pred sistemskimi napakami.

Projektne naloge se bodo neposredno navezoval na realne probleme raziskovalnih projektov, ki jih izvajamo na univerzi.

Ime predmeta: **PODATKOVNI PRAKTIKUM I**
Število ECTS kreditnih točk: **3**

Vsebina:

- Uvod v sintakso in v programiranju v jeziku R.
- Struktura podatkov in obravnava podatkov.
- Osnovne funkcije iz paketa base.
- Učinkovita in ponovljiva obravnava podatkov in statistična analiza v jeziku R.
- Grafični prikaz podatkov s pomočjo osnovnih funkcij in paketa ggplot2.
- Razvoj in objava spletnih aplikacij, ki temeljijo na kodo napisano v jeziku R (s paketom shiny)
- Optimizirano kodiranje v jeziku R: Pogoste napake, ki povzročajo neučinkovitosti kode; Anotacija in kontrola verzij; Preverjanje pravilnosti; Odpravljanje napak; Ocena učinkovitosti; Upravljanje spomina; Računsko intenzivne metode (vključitev kode v jeziku C++ v funkcije, ki so napisane v R s pomočjo paketa Rccp, paralelizacija kode).
- Razvoj programskih paketov v okolju R: Struktura paketov; Generične funkcije in metode; Testiranje in zgradba paketov; Dokumentacija paketov; Kontrola verzij z uporabo GitHuba.
- Aplikacije za analizo podatkov v R: Povzetek statistike; Parametrični in neparametrični testi pomembnosti; Regresija in GLM (linearna, logistična, večkratna); Mešani modeli.

Pri LV bodo uporabljeni realni masovni podatki v okviru raziskovalnih projektov, ki se izvajajo v okviru raziskovalnih programov P1-0285, P1-0404 in P1-0294.

Ime predmeta: **PODATKOVNO INŽENIRSTVO IN DISTRIBUIRANI INFORMACIJSKI SISTEMI**
Število ECTS kreditnih točk: **6**

Vsebina:

Klasične in sodobne raziskovalne teme s področja podatkovnega inženirstva in distribuiranih sistemov. Možne vsebine vključujejo:

- napredna orodja za manipulacijo s podatki, organiziranje podatkov in shranjevanje,
- tehnologije razvoja programske opreme, kot so: računalniško podprto programsko inženirstvo (CASE), orodja za sodelovanje pri razvoju programske opreme (npr. sistemi za nadzor verzij, sistemi za oddajanje nalog),
- podatkovno komuniciranje / računalniška omrežja,
- arhitekture distribuiranih sistemov: klient / strežnika, "peer-to-peer",
- spletne aplikacije in spletne storitve,
- sodobne aplikacije distribuiranih informacijskih sistemov.

Ime predmeta: **SEMINAR IZ PODATKOVNE ZNANOSTI I**
Število ECTS kreditnih točk: **3**

Vsebina:

Pri predmetu bodo študentje preko seminarjev in predavanj gostujočih znanstvenikov iz Slovenije in tujine seznanjeni s posameznimi temami podatkovne znanosti. Poudarek bo na specifičnih raziskavah predavateljev, ki pokrivajo področje podatkovne znanosti.

Študenti bodo v okviru predmeta pripravili tudi seminarsko nalogo, ki bo pokrivala eno od področij zanimanja študenta in se bo navezovala na eno (ali več) izmed predstavljenih tem v okviru predavanj. Seminarji se bodo neposredno navezovali na aplikativne projekte, ki jih izvajamo na univerzi, v okviru katerih se srečujemo s problemom obdelave masovnih podatkov (kot so npr. trenutno potekajoči projekti SUSGRAPE, InnoRenew CoE, ...).

Ime predmeta: **SEMINAR IZ PODATKOVNE ZNANOSTI II**

Število ECTS kreditnih točk: **3**

Vsebina:

Pri predmetu bodo študentje preko seminarjev in predavanj gostujočih znanstvenikov iz Slovenije in tujine seznanjeni s posameznimi temami podatkovne znanosti. Poudarek bo na specifičnih raziskavah predavateljev, ki pokrivajo področje podatkovne znanosti.

Študenti bodo v okviru predmeta pripravili tudi seminarsko nalogo, ki bo pokrivala eno od področij zanimanja študenta in se bo navezovala na eno (ali več) izmed predstavljenih tem v okviru predavanj. Seminarji se bodo neposredno navezovali na aplikativne projekte, ki jih izvajamo na univerzi, v okviru katerih se srečujemo s problemom obdelave masovnih podatkov (kot so npr. trenutno potekajoči projekti SUSGRAPE, InnoRenew CoE ...).

Ime predmeta: **STATISTIKA**

Število ECTS kreditnih točk: **6**

Vsebina:

- Osnove verjetnosti
- Slučajne spremenljivke in porazdelitve
- Pričakovana vrednost
- Neodvisnost in pogojevanje
- Aproksimacija porazdelitev
- Osnove opisne statistike
- Grafične predstavitve podatkov
- Populacijski in vzorčni podatki
- Zanesljivost vzorčnih ocen
- Statistični modeli in njihova vloga v podatkovni znanosti
- Osnove inferenčne statistike, ocenjevanje parametrov.
- Osnove preizkušanja domnev, p-vrednosti, mere prileganja
- Regresijske metode, linerna regresija
- Bayesova paradigma, osnove Bayesove statistike

Ime predmeta: **MATEMATIČNI PRAKTIKUM**

Število ECTS kreditnih točk: **6**

Vsebina:

Pokrite bodo vsebine, ki spadajo med spodaj naštet:

1. Wolfram Mathematica:

- osnove programa, elementarni izračuni, grafi funkcij,
- reševanje standardnih problemov iz analize, linearne algebre, diferencialnih enačb itd.,
- risanje (eksplicitne, implicitne, parametrične prezentacije objektov),
- ustvarjanje interaktivnih in dinamičnih risb,
- grafična predstavitev rešitev NDE in PDE,
- izbrana poglavja.

2. MATLAB oziroma Octave:

- osnove programa,
- vgrajene funkcije,
- delo z matrikami,
- uvažanje in izvažanje podatkov med MATLABom/Octavom in ostalimi formati,
- vizualizacija (risanje različnih objektov),
- programiranje (pisanje funkcij),

- obravnava napak,
- Orodja v MATLABu,
- reševanje realnih problemov z MATLABom ali Octavom.

3. Blender:

- osnove programa,
- spoznavanje uporabniškega vmesnika,
- 3D modeliranje,
- osnove animacije,
- pretvorbe v video.

Ime predmeta: **PODATKOVNI PRAKTIKUM II**

Število ECTS kreditnih točk: **6**

Vsebina:

Faze tipičnega cikla v projektih podatkovnega rudarjenja:

1. Uvoz podatkov,
2. Čiščenje podatkov,

ponavlja v zanki:

3. Preoblikovanje podatkov,
4. Vizualizacija/grafični prikaz podatkov,
5. Modeliranje podatkov,

konec zanke

6. Predstavitev rezultatov analize.

Pri predmetu bo nadgrajena snov iz predmeta *Podatkovni praktikum I*. Predvsem bo poudarek na fazah 2., 3. in 4.

Poleg programskega jezika R, bo predstavljen tudi programski jezik Python.

Predstavljeno bo čiščenje in preoblikovanje podatkov z uporabo naprednih urejevalnikov besedil ter v programskih jezikih R in Python.

Predstavljena bo uporaba tipičnih razširitvenih paketov, ki se najpogosteje uporabljajo v podatkovni znanosti: tidyrr, ggplot2 (R); Numpy, Pandas, Matplotlib (Python). Predstavljena bo tudi uporaba okolij IPython in Jupyter ter napredna uporaba orodja RStudio.

Pri LV bodo uporabljeni realni masovni podatki v okviru raziskovalnih projektov, ki se izvajajo v okviru raziskovalnih programov P1-0285, P1-0404 in P1-0294.

Ime predmeta: **RUDARJENJE MASOVNIH PODATKOV**

Število ECTS kreditnih točk: **6**

Vsebina:

Vsebina predmeta je razdeljena na 15 modulov, od katerih jih 5 sestavlja jedro predmeta, ostalih 10 pa je opcijskih. Izmed opcijskih modulov se vsako leto izvede 5 modulov glede na trenutno najaktualnejše raziskovalne trende.

- Vzporedna in porazdeljena obdelava masovnih podatkov [jedro]
- Analiza povezav [jedro]
- Zgoščevanje na podlagi lokalnosti [jedro]
- Grafi socialnih omrežij [jedro]
- Algoritmi za pretočne podatke [jedro]
- Mere razdalje in najbližji sosed
- Analiza pogostih množic postavk
- Priporočilni sistemi
- Zmanjševanje dimenzij
- Razvrščanje v skupine
- Računalniško oglaševanje

- Strojno učenje
- Napredna vzporedna in porazdeljena obdelava masovnih podatkov
- Napredno zgoščevanje na podlagi lokalnosti
- Napredna analiza povezav

Projektna naloga se bodo neposredno navezovala na realne probleme raziskovalnih projektov, ki jih izvajamo na univerzi.

Ime predmeta: **PRAKTIČNO USPOSABLJANJE**

Število ECTS kreditnih točk: **6**

Vsebina:

Študentje bodo praktično usposabljanje opravljali pod strokovnim mentorstvom v različnih podjetjih, ki iz različnih razlogov v okviru svojega delovanja analizirajo večje količine podatkov npr. za pomoč pri svojem delovanju z vidika odločanja, optimizacije poslovanja, razumevanja potreb in želja uporabnikov, ter v organizacijah, ki razvijajo nove pristope za analizo masovnih podatkov in razvijajo inovativne produkte na osnovi obdelave podatkov.

Študentje bodo lahko prakso opravljali v delovnem okolju oz. v organizacijah, kot so: podjetja, javni zavodi, društva in drugo. Opravljanje prakse je možno tudi v znanstveno-raziskovalnih institucijah članic Univerze na Primorskem.

Praktično usposabljanje bo, v dogovoru s koordinatorjem študijske prakse na oddelku fakultete in ustrezno usposobljenimi mentorji v organizacijah, organizirano tako, da se bodo študentje seznanili z delom in organizacijo v podjetju, ob upoštevanju specifičnega področja delovanja posameznega podjetja.

Ime predmeta: **MAGISTRSKO DELO**

Število ECTS kreditnih točk: **24**

Vsebina:

Celovita obravnava vsebine izbranega predmetnega področja študijskega programa druge stopnje in/ali interdisciplinarna povezava z drugimi predmetnimi področji.

Študenti izberejo temo glede na njihove interese in v dogovoru z mentorjem, predmetnim področjem, s katerim se želijo podrobneje ukvarjati.

Dolžina naloge je točno navedena v fakultetnem pravilniku, ki ureja to področje.

Z izdelavo magistrske naloge študent razvije preiskovalne metode: ustreznost teme, razvoj delovnega načrta, oblikovanje ciljev in hipotez, dokumentarnega in bibliografskega iskanja, vzpostavitev strukture raziskave itd. in se izkaže poznavanje, izbira in aplikacija ustreznih teoretskih podlag.

IZBIRNI PREDMETI ŠTUDIJSKEGA PROGRAMA

Ime predmeta: **VARNOST**

Število ECTS kreditnih točk: **3**

Vsebina:

Varnost operacijskega sistema

- Načrtovanje operacijskega sistema
- Pooblastila za dostop do datotečnega sistema
- Shranjevanje uporabniških gesel

Zaščita na ravni omrežja

- Napad man in the middle

- (Porazdeljena) zavrnitev storitve
- Težave z Wifi (ključ WEP itd.)
- Požarni zid: iptables
- X509 Certifikati

Kriptografske rešitve

- Simetrični sistemi šifriranja
- Asimetrični sistemi šifriranja
- Primeri, kot so AES, DES, RSA, Diffie Hellman, MD5 ...

Orodja za revizijo sistema

- Wireshark, nessus, itd ...

Neobvezno

- LDAP
- Kerberos IPsec

LV se bodo neposredno navezovala na uporabo realnih podatkov v okviru raziskovalnih projektov, ki se izvajajo v okviru raziskovalnega programa P1-0404.

Ime predmeta: **ZBIRANJE IN INTEGRACIJA PODATKOV S SENZORJEV**

Število ECTS kreditnih točk: **3**

Vsebina:

- Zbiranje podatkov: Okolja za množično zbiranje in obdelavo podatkov (Hadoop, Kafka itd.)
- Protokoli in paradigmi: Komunikacija s senzorji in med njimi (protokoli nizke ravni in jeziki)
- Integracija podatkov: Domensko specifični vidiki senzorskih podatkov (semantično modeliranje in integracija / ponovna uporaba ontologije)
- Izbirno: mediacijske rešitve za probleme heterogenosti podatkov

Pri LV bodo uporabljeni realni podatki v okviru raziskovalnih projektov, ki se izvajajo v okviru raziskovalnega programa P1-0404.

Ime predmeta: **RAČUNALNIŠKE DRUŽBENE VEDE**

Število ECTS kreditnih točk: **6**

Vsebina:

Pri predmetu bodo obravnavana poglavja, ki lahko med drugim vključujejo (a niso omejena le na) naslednje teme glede na potrebe in razvoj širšega področja predmeta:

- Računstvo in družbene vede
- Sociološki in psihološki modeli
- Pridobivanje podatkov
- Uporabniške študije
- Podatki s spleta
- Storitve v oblaku
- Prediktivno računsko modeliranje
- Rudarjenje podatkov
- Uporaba metod strojnega učenja
- Analiza družbenih omrežij
- Simulacije

Ime predmeta: **IZBRANA POGlavJA IZ OBDELAVE PODATKOV**

Število ECTS kreditnih točk: **6**

Vsebina:

Pri predmetu bodo obravnavana poglavja, ki lahko med drugim vključujejo (a niso omejena na) naslednje teme glede na potrebe in razvoj področja predmeta:

- klasična poglavja obdelave podatkov (potrjevanje, razvrščanje, povzemanje, združevanje, analiza, razvrščanje).
- novi pristopi obdelave podatkov, ki izhajajo iz najnovejših pristopov pridobivanja podatkov

Predmet je namenjen predstavitvi izbranih tem, ki so zanimive zaradi novih teoretičnih odkritij in zaradi njihove uporabe pri specifičnih aplikacijah. Podrobna vsebina in literatura se določi vsako leto posebej glede na usmeritev predavatelja in potrebe študenta.

Ime predmeta: **IZBRANA POGLAVJA IZ ZAJEMA PODATKOV**

Število ECTS kreditnih točk: **6**

Vsebina:

Pri predmetu bodo obravnavana poglavja, ki lahko med drugim vključujejo (a niso omejena na) naslednje teme glede na potrebe in razvoj področja predmeta:

- klasični postopki zajema podatkov
- novi pristopi z različnimi viri podatkov (fizični, fiziološki, digitalni, človeški in drugi).

Predmet je namenjen predstavitvi izbranih tem, ki so zanimive zaradi novih teoretičnih odkritij in zaradi njihove uporabe pri specifičnih aplikacijah. Podrobna vsebina in literatura se določi vsako leto posebej glede na usmeritev predavatelja in potrebe študenta.

Ime predmeta: **IZBRANA POGLAVJA IZ VIZUALIZACIJE PODATKOV II**

Število ECTS kreditnih točk: **6**

Vsebina:

Pri predmetu bodo obravnavana poglavja, ki lahko med drugim vključujejo (a niso omejena na) naslednje teme glede na potrebe in razvoj področja predmeta:

- klasične pristopi vizualizacije podatkov (oblikovanje in ustvarjanje lahko sporočljivih in razumljivih grafičnih ali vizualnih predstavitev velike količine kompleksnih kvantitativnih in kvalitativnih podatkov).
- novi in napredni pristopi vizualizacije podatkov, ki temeljijo na novih vrstah podatkov in možnostih vizualizacije (npr. nove modalitete)

Predmet je namenjen predstavitvi izbranih tem, ki so zanimive zaradi novih teoretičnih odkritij in zaradi njihove uporabe pri specifičnih aplikacijah. Podrobna vsebina in literatura se določi vsako leto posebej glede na usmeritev predavatelja in potrebe študenta.

Ime predmeta: **TEMELJI PODATKOVNE ZNANOSTI IN UMETNE INTELIGENCE**

Število ECTS kreditnih točk: **6**

Vsebina:

Študenti se bodo seznanili z matematičnimi osnovami naslednjih tem:

- Visokodimenzijski prostor: zakon velikih števil, geometrija visokih dimenzij, Gaussovi v visokih dimenzijah
- Učno ogrodje PAC: učni model PAC, končni nizi hipotez
- Treniranje proti testiranju: teorija posploševanja, dimenzija VC, mere kompleksnosti
- Modeli učenja: linearni model, nelinearna transformacija, overfitting in validacija, principi učenja
- Naključni sprehodi in Markovske verige: Konvergenca naključnih sprehodov, Naključni sprehodi v evklidskem prostoru, Monte Carlo Markove verige
- Markovljev proces odločanja in krepitveno učenje: Bellmannove enačbe, dinamično programiranje, učenje Monte Carlo
- Učenje na strukturi grafov: model graf nevronske mreže (GNN), GNN in konvolucije grafov, GNN in verjetnostni modeli grafa, GNN in problemi teorije grafov

- Agentsko modeliranje in teorija iger: kooperativna teorija iger in modeli pravičnosti, načrtovanje algoritemskih mehanizmov

Ime predmeta: **MASOVNI PODATKI V BIOINFORMATIKI**

Število ECTS kreditnih točk: **6**

Vsebina:

Zaradi izboljšav eksperimentalne tehnologije za področju strukturne biologije, se je v zadnji 10 letih zelo povečalo število rešenih 3D struktur proteinov, ki pa so ključni za razlago molekularnih mehanizmov. Pri predmetu bodo obravnavana poglavja, ki lahko med drugim vključujejo (a niso omejena le na) naslednje teme glede na potrebe in razvoj širšega področja predmeta:

- spoznavanje podatkovnih baz na področju strukturne bio-informatike: PDB (Protein Data Bank), EMDB (the Electron Microscopy Data Bank), PorBis, SCOPe, AlphaFold DB
- spoznavanje orodij in knjižnic za obdelavo masovnih podatkov
- Osnove načrtovanja zdravilnih učinkovin
- Osnove intermolekularne prepoznave
- Spoznavanje s kemijskimi virtualnimi knjižnicami
- Kemijski big-data in analiza kemijskega prostora
- Kemoinformacijski aplikativni problemi

Projektne naloge se bodo neposredno navezovala na realne probleme iz bioinformatike.

Podrobna vsebina in literatura se določi vsako leto posebej glede na usmeritev predavatelja in potrebe študenta.

Ime predmeta: **GRAFI ZNANJA**

Število ECTS kreditnih točk: **6**

Vsebina:

Predmet bo obravnaval teme, ki lahko vključujejo (vendar niso omejene na) naslednje, glede na potrebe in razvoj obravnavanih vsebin:

- Osnove: Podatki vs. Znanje; Predstavitev na osnovi grafov;
- Logične osnove: Predikatni račun; Opisna logika (OL); Sklepanje na osnovi pravil (Datalog, veriženje);
- Jeziki za grafe znanja: RDF, RDFS; SPARQL; OWL;
- Sklepanje: Veriženje naprej/nazaj; Profili sklepanja v OWL (RL, EL, QL); Sklepanje s pravili vs. sklepanje v OL;
- Sistemi: Baze trojic (npr. GraphDB); Sklepanje (Pellet, HermiT, FaCT++); Integracija KG–DB;
- Grafi znanja iz prakse: YAGO, OpenAlex, Wikidata;
- Uporabe: Iskanje (Google KG); KG in LLM (poravnava grafov znanja); Integracija podatkov;
- Projekt / praktično delo.

Projektne naloge bodo neposredno povezane z realnimi problemi raziskovalnih projektov, ki potekajo na univerzi.