

UČNI NAČRT PREDMETA / COURSE SYLLABUS (leto / year 2017/18)						
<b>Predmet:</b>		Uvod v bioinformatiko				
<b>Course title:</b>		Introduction to bioinformatics				
<b>Študijski program in stopnja</b> Study programme and level		<b>Študijska smer</b> Study field		<b>Letnik</b> Academic year	<b>Semester</b> Semester	
Interdisciplinarni magistrski študijski program Računalništvo in matematika		ni smeri		1 ali 2	prvi	
Interdisciplinary Master's study programme Computer Science and Mathematics		none		1 or 2	first	
<b>Vrsta predmeta / Course type</b>				izbirni / elective		
<b>Univerzitetna koda predmeta / University course code:</b>				63520		
<b>Predavanja</b> Lectures	<b>Seminar</b> Seminar	<b>Vaje</b> Tutorial	<b>Klinične vaje</b> work	<b>Druge oblike študija</b>	<b>Samost. delo</b> Individ. work	<b>ECTS</b>
45	20	10			105	6
<b>Nosilec predmeta / Lecturer:</b>		prof. dr. Blaž Zupan				
<b>Jeziki / Languages:</b>		<b>Predavanja / Lectures:</b>				
		<b>Vaje / Tutorial:</b>				
<b>Pogoji za vključitev v delo oz. za opravljanje študijskih obveznosti:</b>				<b>Prerequisites:</b>		
Vpis v letnik študija.				Enrolment in the programme.		
<b>Vsebina:</b>				<b>Content (Syllabus outline):</b>		

Kako podobna so si med seboj živa bitja? Smo ljudje res potomci neandertalcev? Kako so se živalske vrste lahko prilagodile na življenjska okolja? Kateri geni so odgovorni za pojav različnih bolezni? Zakaj vsako leto potrebujemo drugo cepivo proti gripi?

Današnja biologija je polna zanimivih vprašanj in nikoli prej nismo bili tako blizu odgovorom. Nedavno razvite eksperimentalne biotehnologije omogočajo pridobivanje velikih količin eksperimentalnih podatkov: o genomih različnih vrst in osebkov, o genskih izrazih, koncentracij proteinov, vplivih učinkovin na delovanje celice, ipd. Tovrstni podatki so danes dostopni v javnih spletnih podatkovnih bazah in jih je potrebno statistično in matematično obdelati, v njih poiskati skrite vzorce in jih primeren način prikazati. Temu so namenjena orodja bioinformatike, področja, brez katerega danes ne bi mogli več odgovoriti niti na eno od zgoraj zapisanih vprašanj.

Gre za interdisciplinarno področje, ki združuje metode iz statistike, matematike, vizualizacije podatkov, strojnega učenja in umetne inteligence.

Pri predmetu si bomo v teoriji in na praktičnih primerih ogledali osnovne bioinformatične metode ter se spoznali s sledečimi vsebinami:

1. Osnove celične biologije
2. Statistične lastnosti nukleotidnih zaporedij
3. Računske tehnike za iskanje genov v nukleotidnih zaporedjih
4. Tehnike poravnave zaporedij, algoritem BLAST
5. Verjetnostni modeli zaporedij, markovske

How similar are living organisms? Have human indeed descended from Neanderthals? How did various species adapt to living environments? Which genes are responsible for susceptibility to various diseases? Why we need a different flu vaccine each year?

Modern biology poses many interesting questions, and never before have we been so close to answering them. Recently developed experimental biotechnologies allow us to gather vast amounts of experimental data. From genomes of various species, including that of *H. sapiens*, to gene expression, protein concentrations, effects of various chemicals to cell processes, and similar. Vast number of experimental data sets is today available in open, public repositories, and requires further statistical and mathematical analysis to discover useful and applicable patterns. The methods and techniques for such analysis is developed within the field of bioinformatics, which combines techniques from statistics, computer science, mathematics, data mining and visualization, machine learning and artificial intelligence. During the course, the students will in theory and practice get familiar with the following topics:

1. Basics of molecular biology
2. Statistical properties of nucleotide sequences
3. Computational approaches to gene finding and annotation
4. Sequence alignment (BLAST)
5. Probabilistic models for nucleotide sequences, Markov chain models
6. Computational techniques for assessment of genetic distances between species and

<p>verige</p> <p>6. Računske tehnike ocenjevanja genskih razlik med predstavniki osebkov iste vrste in osebkov različnih vrst</p> <p>7. Filogenetska analiza, računski pristopi k odkrivanju evolucijskih dreves</p> <p>8. Računske primerjave genomov</p> <p>9. Analiza podatkov o genskih izrazih, uporaba tehnik uvrščanja in razvrščanja v skupine, genski izrazi v medicinski diagnostiki in prognostiki, analiza obogatenosti genskih skupin, vizualizacijske tehnike, genske mreže</p> <p>10. Integrativna bioinformatika: uporaba različnih baz podatkov in baz znanj v namene odkrivanja smiselnih vzorcev v biomedicinskih podatkih</p> <p>Teoretično predstavitev računskih pristopov in tehnik bo spremljal pregled javno dostopnih baz podatkov s področja, prikaz delovanja ustrezne odprtokodne programske opreme in prikaz uporabe tehnik in orodij pri reševanju praktičnih problemov s področja biomedicine in sistemske biologije. Pri analizi podatkov bomo uporabljali moderna skriptna okolja (npr. Python) in že razvite bioinformatične knjižnice (npr. Biopython in Orange). Uvod v uporabo bioinformatične programske opreme bo podan na predavanjih, praktično pa bomo ta orodja spoznali na vajah in pri projektne delu.</p>	<p>individuals within the same species</p> <p>7. Phylogenetic analysis, computational techniques for construction of evolution trees</p> <p>8. Computational comparison of genomes</p> <p>9. Analysis of transcriptome, utility of data mining and visualization techniques, gene set enrichment analysis, gene networks, applications in biomedicine</p> <p>10. Integrative bioinformatics: how to combine various data sources and various modeling techniques to discover patterns in biomedical data sets</p> <p>Theoretical study of the above concepts will be accompanied with familiarization with public data repositories and open-source tools to assess the data and perform subsequent analysis. We will use scripting tools (e.g. Python) and already developed bioinformatics libraries (e.g., Biopython and Orange).</p>
--	--

**Temeljni literatura in viri / Readings:**

1. N. Cristianini, M. W. Hahn (2007) Introduction to computational genomics: A case study approach, Cambridge University Press, Cambridge, UK.

2. R. Durbin, S. Eddy, A. Krogh, G. Mitchison (2010) Biological sequence analysis, Cambridge University Press.

#### **Cilji in kompetence:**

Cilj predmeta je študente seznaniti z osnovnimi računskimi tehnikami, orodji in prosto dostopnimi bazami podatkov s področja bioinformatike. V okviru predmeta bodo predstavljene osnove biologije in genomike, ki bodo študentom računalništva omogočale razumevanje problemske domene tako, da lahko nato s pomočjo matematičnih, statističnih in računskih pristopov, ki jih bo študent spoznal pri predmetu, poišče odgovore na sicer kompleksna vprašanja s področij evolucije in razvoja živih bitij, povezav med geni in biološkimi procesi, vpliv genskih predispozicij na razvoj bolezni, in podobnih.

#### **Objectives and competences:**

This is an introductory course to bioinformatics. During the course the students will become familiar with computational methods and tools that can be used in bioinformatics, and with publically available data bases in molecular biology. The course will start with introduction to molecular biology and genomics, which will allow students of computer science to apply mathematical, statistical and computational techniques to problems from evolution of living organisms, interactions of genes and biological processes, interactions between genome and phenotypes and diseases, and similar.

#### **Predvideni študijski rezultati:**

Znanje in razumevanje:  
Poznavanje metod bioinformatike in osnovnih orodij za bioinformatično analizo podatkov. Poznavanje osnovnih bioinformatičnih podatkovnih baz. Poznavanje načinov gradnje sistemov za bioinformatično analizo iz obstoječih analitičnih komponent.

#### **Uporaba:**

Uporaba bioinformatičnih tehnik v namene analize podatkov v biomedicini, genomiki in sistemski biologiji.

#### **Refleksija:**

#### **Intended learning outcomes:**

Knowledge and understanding:  
Knowledge of methods in bioinformatics and computational techniques for analysis of data from molecular biology. Familiarization with principal bioinformatics data source and data repositories. Engineering skills for development of bioinformatics systems from a set of existing data analysis components.

#### **Application:**

Application of bioinformatics in biomedicine, genomics and systems biology.

#### **Reflection:**

Razumevanje primernosti teoretičnih metod za reševanje praktičnih problemov ter njihovih omejitev, sposobnost analitičnega razmišljanja, sposobnost analize in reševanja kompleksnih praktičnih problemov s področja sodobne biomedicine z razvojem računskih analitičnih postopkov.

Prenosljive spretnosti - niso vezane le na en

predmet:

Kombiniranje znanj pridobljenih pri predmetih s področja programiranja, statistike, strojnega učenja in odkrivanja znanj iz podatkov. Spretnosti iskanja in uporabe domače in tuje literature, uporaba primerne (predvsem odprtokodne) programske opreme, identifikacija in reševanje kompleksnih problemov.

Which are appropriate practical applications of theoretical methods of bioinformatics data analysis? What are their limitations? How can bioinformatic data analysis systems be used in practice to solve complex problems?

Transferable skills:

Students will be able to combine the knowledge from other courses that cover programming, statistics, machine learning and data mining. The course will require students to acquire skills in literature search and search for existing algorithmic solutions and code snippets, and engineering skills for solving real-life complex problems.

#### Metode poučevanja in učenja:

Predavanja s podporo avdio-vizualne opreme, sprotni razvoj programskih rešitev, laboratorijske vaje v računalniški učilnici z ustrežno programsko opremo. Delo posamezno in v skupinah. Velik poudarek na praktičnem delu (npr. razvoj skript za pregledovanje in analizo podatkov) in reševanju praktičnih problemov.

#### Learning and teaching methods:

Combined lecturing with simultaneous use of the blackboard and computer projection (coding, visualization of models, results). Lab work in computer-equipped lecture rooms. Individual and work in team. Emphasis on practical problem solving.

#### Načini ocenjevanja:

Način (pisni izpit, ustno izpraševanje, naloge, projekt):

Delež (v %) /

Weight (in %)

#### Assessment:

Type (examination,

50%

<p>Sprotno preverjanje (domače naloge, kolokviji in projektno delo)</p> <p>Končno preverjanje (pisni in ustni izpit)</p> <p>Ocene: 6-10 pozitivno, 1-5 negativno</p> <p>(v skladu s Statutom UL)</p>	<p>50%</p>	<p>oral, coursework, project):</p> <p>Continuing (homework, midterm exams, project work)</p> <p>Final (written and oral exam)</p> <p>Grading: 6-10 pass, 1-5 fail (according to the rules of University of Ljubljana)</p>
--	------------	---

**Reference nosilca / Lecturer's references:**

BELLAZZI, Riccardo, ZUPAN, Blaž. Predictive data mining in clinical medicine : current issues and guidelines. International journal of medical informatics, ISSN 1386-5056. [Print ed.], 2008, vol. 77, no. 2, str. 81-97, ilustr. [COBISS.SI-ID 6280788]

MRAMOR, Minca, LEBAN, Gregor, DEMŠAR, Janez, ZUPAN, Blaž. Visualization-based cancer microarray data classification analysis. Bioinformatics, ISSN 1367-4803. [Print ed.], 2007, vol. 23, no. 16, str. 2147-2154, ilustr. [COBISS.SI-ID 6087252]

VAN DRIESSCHE, Nancy, DEMŠAR, Janez, BOOTH, Egzi O., HILL, Paul, JUVAN, Peter, ZUPAN, Blaž, KUSPA, Adam, SHAULSKY, Gad. Epistasis analysis with global transcriptional phenotypes. Nature genetics, ISSN 1061-4036, May 2005, vol. 37, no. 5, str. 471-477, ilustr. [COBISS.SI-ID 4712532]

ZUPAN, Blaž, DEMŠAR, Janez, BRATKO, Ivan, JUVAN, Peter, HALTER, John A., KUSPA, Adam, SHAULSKY, Gad. GenePath : a system for automated construction of genetic networks from mutant data. Bioinformatics, ISSN 1367-4803. [Print ed.], 2003, vol. 19, no. 3, str. 383-389. [COBISS.SI-ID 3415124]

ZUPAN, Blaž, BOHANEC, Marko, DEMŠAR, Janez, BRATKO, Ivan. Learning by discovery concept hierarchies. Artificial intelligence, ISSN 0004-3702. [Print ed.], 1999, vol. 109, str. 211-242. [COBISS.SI-ID 14228007]