

UČNI NAČRT PREDMETA / COURSE SYLLABUS

Predmet:	Podatkovno in tekstovno rudarjenje
Course title:	Data and Text Mining

Študijski program in stopnja Study programme and level	Modul Module	Letnik Academic year	Semester Semester
Informacijske in komunikacijske tehnologije, 2. stopnja	Tehnologije znanja	1	1
Information and Communication Technologies, 2 nd cycle	Knowledge Technologies	1	1

Vrsta predmeta / Course type Izbirni / Elective

Univerzitetna koda predmeta / University course code: IKT2-713

Predavanja Lectures	Seminar Seminar	Sem. vaje Tutorial	Lab. vaje Laboratory work	Druge oblike	Samost. delo Individ. work	ECTS
60	30			60	450	20

**Navedena porazdelitev ur velja, če je vpisanih vsaj 15 študentov. Drugače se obseg izvedbe kontaktnih ur sorazmerno zmanjša in prenese v samostojno delo. / This distribution of hours is valid if at least 15 students are enrolled. Otherwise the contact hours are linearly reduced and transferred to individual work.*

Nosilec predmeta / Lecturer: Prof. dr. Nada Lavrač
Prof. dr. Dunja Mladenić
Prof. dr. Bojan Cestnik
Doc. dr. Martin Žnidaršič

Jeziki / Languages: **Predavanja / Lectures:** slovenščina, angleščina / Slovenian, English
Vaje / Tutorial:

Pogoji za vključitev v delo oz. za opravljanje študijskih obveznosti:

Zaključen študijski program prve stopnje s področja naravoslovja, tehnike ali računalništva.

Prerequisites:

Student must complete first-cycle study programmes in natural sciences, technical disciplines or computer science.

Vsebina:

Uvod:
uvod v podatkovno rudarjenje in odkrivanje zakonitosti v podatkih, povezava s strojnim učenjem, vizualizacija podatkov, vzorcev in modelov, predstavitev CRISP-DM metodologije odkrivanja zakonitosti v podatkih ter osnove upravljanja znanja

Content (Syllabus outline):

Introduction:
introduction to knowledge data mining and knowledge discovery in databases, relation with machine learning, visualization of data, patterns and models, presentation of the CRISP-DM knowledge discovery methodology, and the basics of knowledge management

<p>Predstavitev in manipulacija podatkov: predstavitev standardnih oblik zapisa različnih vrst podatkov ter kreiranje in manipulacija tabelarnih podatkov, podatkovnih baz in skladišč ter obravnava tekstovnih, spletnih in večpredstavnih podatkov</p> <p>Tehnike rudarjenja tabelarnih podatkov: predstavitev posameznih tehnik rudarjenja podatkov: predstavitev preiskovalnih heuristik in metod za učenje odločitvenih dreves, učenje klasifikacijskih in povezovalnih pravil, razvrščanje v skupine, odkrivanje podskupin, učenje regresijskih dreves in relacijsko podatkovno rudarjenje</p> <p>Tehnike rudarjenja tekstovnih, spletnih in večpredstavnih podatkov: predstavitev posameznih tehnik za analizo tekstovnih, spletnih in večpredstavnih podatkov ter metod vizualizacije podatkov</p> <p>Ocenjevanje: predstavitev metod za ocenjevanje kvalitete naučenih vzorcev in modelov ter metodologija evalvacije rezultatov</p> <p>Praktično usposabljanje: praktična uporaba izbranih orodij za manipulacijo in rudarjenja podatkov</p>	<p>Data representation and manipulation: presentation of standard data formats, creation and manipulation of tabular data, databases and data warehouses, as well as handling of text, web and multimedia data</p> <p>Techniques for mining of tabular data: presentation of specific data mining techniques: presentation of search heuristics, decision tree learning, learning classification and association rules, clustering, subgroup discovery, regression tree learning, and relational data mining</p> <p>Techniques for mining of text, web and multimedia data: presentation of specific techniques for text, web and multimedia mining, and data visualization</p> <p>Evaluation: presentation of methods for estimating the quality of induced patterns and models, and methodology for result evaluation</p> <p>Practical training: practical use of selected data manipulation and data mining tools</p>
--	--

Temeljna literatura in viri / Readings:

Izbrana poglavja iz naslednjih knjig: / Selected chapters from the following books:

- I. Witten, and F. Eibe, *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann, 1999. ISBN 978-1-558-60552-7
- D. Mladenić, N. Lavrač, M. Bohanec, and S. Moyle, Eds. *Data Mining and Decision Support: Integration and Collaboration*. Kluwer, 2003. ISBN 1-4020-7388-7
- I. Kononenko, and M. Kukar, *Machine Learning and Data Mining*. Horwood Publishing, 2007. ISBN 978-1-904-27521-3
- T. Mitchell, *Machine Learning*. McGraw Hill, 1997. ISBN 978-0-070-42807-2
- M. Berthold, and D. J. Hand, Eds. *Intelligent Data Analysis: An Introduction*. Springer, Berlin-Heidelberg, 1999. ISBN 978-3-540-65808-5
- S. Džeroski, and N. Lavrač, Eds. *Relational Data Mining*. Springer 2001. ISBN 3-540-42289-7
- J. Fürnkranz, D. Gamberger, and N. Lavrač, *Foundations of Rule Learning*. Springer 2012. ISBN 978-3-540-75196-0
- C. Sammut, G. Webb Eds. *Encyclopedia of machine learning and data mining*. Heidelberg [etc.]: Springer. 2016. (selected entries)

- S. Chakrabarti, *Mining the Web: Analysis of Hypertext and Semi Structured Data*, Morgan Kaufmann, 2002. ISBN 1-55860-754-4
- U. Fayyad, G.G. Grinstein, and A. Wierse, Eds. *Information Visualization in Data Mining and Knowledge Discovery*. Morgan Kaufmann. 2001. ISBN 978-1-558-60689-0
- M. Bramer, *Principles of Data Mining*. Springer, 2007. ISBN 978-1-84628-765-7

Cilji in kompetence:

Odkrivanje zakonitosti v podatkih je proces odkrivanja vzorcev in modelov, opisanih s pravili ali drugimi človeku razumljivimi formalizmi za predstavitev znanja. Najpomembnejši del tega procesa predstavlja podatkovno rudarjenje, ki vključuje uporabo metod, tehnik in orodij za avtomatsko odkrivanje vzorcev in konstrukcijo modelov iz podatkov. Cilji predmeta so:

- predstaviti osnove podatkovnega rudarjenja, postopke odkrivanja zakonitosti v podatkih, metodologijo CRISP-DM ter osnove upravljanja znanja
- predstaviti standardne oblike zapisa različnih vrst podatkov, usposobiti študente za manipulacijo tabelarnih podatkov, podatkovnih baz in skladišč ter tekstovnih, spletnih in večpredstavnih podatkov
- predstaviti izbrane metode in tehnike rudarjenja tabelarnih podatkov
- predstaviti izbrane metode in tehnike rudarjenja tekstovnih, spletnih in večpredstavnih podatkov
- usposobiti študente za praktično uporabo izbranih orodij podatkovnega rudarjenja in metod za evalvacijo rezultatov

Objectives and competences:

Knowledge discovery in databases is a process of discovering patterns and models, described by rules or other human understandable representation formalisms. The most important step in this process is data mining, performed by using methods, techniques and tools for automated discovery of patterns and construction of models from data. The course objectives are to:

- introduce the basics of data mining, the process of knowledge discovery in databases, the CRISP-DM methodology and the basics of knowledge management
- present standard data formats, train students for the manipulation of tabular data, databases and data warehouses, as well as text, web and multimedia data
- present selected methods and techniques for mining of tabular data
- present selected methods and techniques for text, web and multimedia mining
- train students for practical use of selected data mining techniques and evaluation methods

Predvideni študijski rezultati:

Študenti bodo z uspešno opravljenimi obveznostmi tega predmeta pridobili:

- sposobnost raziskave, izbire in organizacije informacij kot tudi sinteze rešitev ter predvidevanja njihovih posledic,
- obvladanje strategij in raziskovalnih metod za reševanje problemov in odločanje,
- sposobnost uporabe znanja v praksi,
- postavljanje in doseganje profesionalnih ciljev,
- samostojno, odgovorno in kreativno izvajanje aktivnosti,
- zavezanost profesionalni etiki in regulativi,

Intended learning outcomes:

Students successfully completing this course will acquire:

- ability to research, select and organise information so as to synthesise and anticipate solutions and consequences,
- to adopt strategies and methods appropriate to problem solving and decision making,
- ability to apply the theory in to a practice,
- setting and achieving professional objectives,
- to carry out activities in an autonomous, responsible and creative manner

- sodelovanje z drugimi na skupnih zadolžitvah in problemih.
- poznavanje konceptov in principov rudarjenja podatkov in odkrivanja zakonitosti v podatkovnih bazah,
- zmožnost uporabljanja specifičnih tehnik rudarjenja podatkov,
- zmožnost izdelave aplikacij z orodji za rudarjenje podatkov,
- zmožnost ocenjevanja in evalvacije rezultatov podatkovnega rudarjenja,
- sposobnost izbire in uporabe ustreznih teorij in programskih orodij za analizo večpredstavnih vsebin.

- complying with professional ethics and regulatory body policies,
- to cooperate with others on common tasks and problems
- knowledge of concepts and principles of data mining and knowledge discovery in databases
- ability to use specific data mining techniques
- ability to develop applications using data mining tools
- capacity of evaluating data mining results
- ability to identify and apply appropriate technologies and software tools from the area of multimedia mining

Metode poučevanja in učenja:

Predavanja, seminar, konzultacije, individualno delo

Learning and teaching methods:

Lectures, seminar, consultations, individual work

Načini ocenjevanja:	Delež (v %) / Weight (in %)	Assessment:
Seminar	50 %	Seminar
(pisni ali ustni) izpit	50 %	(written or oral) exam

Reference nosilca / Lecturer's references:

- J. Fürnkranz, D. Gamberger, and **N. Lavrač**, Foundations of Rule Learning. Springer 2012.
- B. Sluban, D. Gamberger, and **N. Lavrač**, Ensemble-based noise detection: noise ranking and visual performance evaluation. *Data Min. Knowl. Discov.* 28(2): 265-303, 2014.
- A. Vavpetič, V. Podpečan, and **N. Lavrač**, Semantic subgroup explanations. *J. Intell. Inf. Syst.* 42(2): 233-254, 2014.
- J. Brank, **D. Mladenić**, M. Grobelnik. Feature construction in text mining. In: C. Sammut, G. Webb Eds. *Encyclopedia of machine learning and data mining*. Heidelberg [etc.]: Springer. 2016.
- I. Petrič, and **B. Cestnik**, Predicting future discoveries from current scientific literature. In: KUMAR, Vinod D. (ur.). *Biomedical Literature Mining, Methods in Molecular Biology*, ISSN 1064-3745, vol. 1159). New York [etc.]: Humana Press, cop. 2014, pp. 159-168.