

UČNI NAČRT PREDMETA / COURSE SYLLABUS

Predmet:	Napredne jezikovne tehnologije
Course title:	Advanced Language Technologies

Študijski program in stopnja Study programme and level	Modul Module	Letnik Academic year	Semester Semester
Informacijske in komunikacijske tehnologije, 3. stopnja	Tehnologije znanja	1	1
Information and Communication Technologies, 3 rd cycle	Knowledge Technologies	1	1

Vrsta predmeta / Course type Izbirni / Elective

Univerzitetna koda predmeta / University course code: IKT3-724

Predavanja Lectures	Seminar Seminar	Sem. vaje Tutorial	Lab. vaje Laboratory work	Druge oblike	Samost. delo Individual work	ECTS
15	15			15	105	5

**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. Tomaž Erjavec

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

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

Zaključen študij druge stopnje s področja informacijskih ali komunikacijskih tehnologij ali zaključen študij druge stopnje na drugih področjih z znanjem osnov s področja predmeta. Potrebna so tudi osnovna znanja matematike, računalništva in informatike.

Prerequisites:

Completed second cycle studies in information or communication technologies or completed second cycle studies in other fields with knowledge of fundamentals in the field of this course. Basic knowledge of mathematics, computer science and informatics is also requested.

Vsebina:

Uvod:
Razvoj jezikoslovja in računalniškega jezikoslovja, kompleksnost jezika, ravni analize jezika, pregled aplikacij in metod.

Analiza jezika z metodami strojnega učenja:
Relevantne metode strojnega učenja, primeri uporabe za avtomatizirano označevanje na morfološki, sintaktični in semantični ravni.

Standardi za zapis:
Zgodovina standardizacije, kodni sistemi, XML, Text Encoding Initiative, ISO, metode

Content (Syllabus outline):

Introduction:
Development of linguistics and computational linguistics, complexity of language, levels of linguistic analysis, overview of applications and methods.

Text analysis with machine learning methods:
Relevant methods of machine learning, use cases: automatic morphological, syntactic and semantic annotation.

Encoding standards:
History of standardisation, coding of characters,

evalvacije.

Jezikoslovne raziskovalne infrastrukture:
Odporna znanost, digitalna humanistika, pravne
in etične dimenzije ravnanja z besedili,
raziskovalna infrastruktura CLARIN.

XML, Text Encoding Initiative, ISO, evaluation
methods.

Research infrastructures for linguistics:
Open science, Digital humanities, ethical and
legal considerations of dealing with language
data, CLARIN research infrastructure.

Temeljni literatura in viri / Readings:

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

- D. Jurafsky, and J.H. Martin. *Speech and Language Processing, 2nd Edition*. Prentice-Hall, 2008. ISBN 978-0131873216
- R. Mitkov (ed.). *The Oxford Handbook of Computational Linguistics*. Oxford University Press, 2003. ISBN 978-0-19-823882-9
- C. Manning, and H. Schütze. *Foundations of Statistical Natural Language Processing*. MIT Press. 1999. ISBN 0-262-13360-1
- N. Ide and J. Pustejovsky (eds.). *Handbook of Linguistic Annotation*. Springer. 2017. ISBN 978-94-024-0881-2

Cilji in kompetence:

Jezikovne tehnologije zajemajo metode in aplikacije obdelave naravnega jezika na računalniku.

Slušatelji pridobijo teoretično razumevanje in praktične izkušnje s področij jezikovnih tehnologij in računalniškega jezikoslovja, kar je predpogoj za učinkovito delo na računalniški obdelavi jezikovnih podatkov.

Cilji predmeta so (a) predstaviti osnove jezikovnih tehnologij, (b) predstaviti zapis in označevanje jezikovnih virov in (c) izbrane metode in tehnike jezikovnih tehnologij.
Poudarek predmeta je na obravnavi slovenskega jezika.

Študenti bodo obvladali osnove jezikovnih tehnologij in bodo usposobljeni za praktično uporabo izbranih metod in orodij.

Objectives and competences:

Language technologies comprise methods and applications of computer processing of natural language.

Students will gain basic theoretical understanding and practical experience of language technologies and computational linguistics, which is a prerequisite for effective work on computer processing of language data.

The course objectives are to (a) introduce the basics of language technologies, (b) present the coding and annotation of language resources, and (c) present selected methodologies and techniques used in language technologies. The focus of the course is on the processing of Slovene language.

The students will master the basics of language technologies and will be capable of using selected methods and tools in practice.

Predvideni študijski rezultati:

Obvladana uporaba izbranih metod in tehnik jezikovnih tehnologij, usposobljenost za praktično uporabo izbranih metod in orodij.

Intended learning outcomes:

Mastering of selected methods and techniques of language technologies, capability of practical use of selected methods and techniques.

Metode poučevanja in učenja:

Predavanja, seminar, konzultacije, samostojno delo

Learning and teaching methods:

Lectures, seminar, consultancy, individual work

Načini ocenjevanja:	Delež (v %) / Weight (in %)	Assessment:
Pisni ali ustni izpit	50 %	Written or oral exam
Seminarska naloga	25 %	Seminar work
Ustni zagovor seminarske naloge	25 %	Oral defense of the seminar work

Reference nosilca / Lecturer's references:

- Y. Scherrer, **T. Erjavec**. Modernising historical Slovene words. *Natural language engineering*, ISSN 1351-3249, 2016, vol. 22, no. 6, str. 881-905.
- **T. Erjavec**, The IMP historical Slovene language resources. *Language resources and evaluation*, 23 str., doi: 10.1007/s10579-015-9294-7, 2015.
- **T. Erjavec**, MULTTEXT-East. V: Ide, N. (ur.), Pustejovsky, J. (ur.), *Handbook of linguistic annotation*. Amsterdam: Springer. 2017, str. 441-462.
- D. Divjak, S. Sharoff, **T. Erjavec**. Slavic corpus and computational linguistics. *Journal of Slavic linguistics*, ISSN 1068-2090, 2017, vol. 25, no. 2, str. 171-198, doi: 10.1353/jsl.2017.0008
- **T. Erjavec**, N. Ljubešić, N. Logar. The slWaC corpus of the Slovene Web. *Informatica : an international journal of computing and informatics*, ISSN 0350-5596, Mar. 2015, vol. 39, no. 1, str. 35-42