



STUDIJŲ DALYKO (MODULIO) APRAŠAS

Dalyko (modulio) pavadinimas	Kodas
Natūralios kalbos apdorojimas	

Dėstytojas (-ai)	Padalinys (-iai)
Koordinuojantis: dr. Gražina Korvel Kitas (-i): dr. Gintautas Tamulevičius	Matematikos ir informatikos fakultetas Duomenų mokslo ir skaitmeninių technologijų institutas

Studijų pakopa	Dalyko (modulio) tipas
Pirmoji	Pasirenkamasis

Igyvendinimo forma	Vykdyto laikotarpis	Vykdyto kalba (-os)
Auditorinė	7 semestras	Lietuvių

Reikalavimai studijuojančiajam	
Išankstiniai reikalavimai: Matematika informacinėms sistemoms, Statistiniai duomenų analizės metodai.	Gretutiniai reikalavimai (jei yra):

Dalyko (modulio) apimtis kreditais	Visas studento darbo krūvis	Kontaktinio darbo valandos	Savarankiško darbo valandos
5	133	64	69

Dalyko (modulio) tikslas: studijų programos ugdomos kompetencijos

Dalyko tikslas – siekiama, kad studentai susipažintų su pagrindiniais natūralios kalbos apdorojimo principais ir teoriniais aspektais, ugdytų gebėjimus taikyti natūralios kalbos apdorojimo metodus sprendžiant praktinius uždavinius. Studentų dėmesys sutelkiamas į kompiuterio panaudojimą natūralios kalbos procesams valdyti. Modulis apima gramatinę kalbos nagrinėjimą, sintaksinę analizę, semantinę interpretaciją ir kalbos pragmatiką. Apžvelgiami metodai ir jų naudojimo sritys. Studentų atliekami laboratoriniai darbai atspindi pagrindines natūralios kalbos apdorojimo dalyko temas. Projektinė užduotis leidžia giliau įvertinti atskirą kurso temą.

Dalyko (modulio) studijų siekiniai	Studijų metodai	Vertinimo metodai
Gebės natūralios kalbos apdorojimo užduotis atlikti tiek savarankiškai, tiek dirbant komandoje.	Individualūs ir grupiniai laboratoriniai darbai, diskusijos, projektinė užduotis, darbų pristatymas	Laboratorinių darbų ataskaitos, kontrolinis darbas, projektinės užduoties atlikimas ir pristatymas, egzaminas.
Gebės atlikti tinkamos literatūros, susijusios su natūralios kalbos apdorojimo algoritmų taikymu paiešką ir analizę, įsisavinti naujas žinias ir jas pritaikyti praktikoje.	Literatūros analizė, projektinė užduotis	
Gebės atlikti natūralios kalbos apdorojimo tyrimus, formuluoti tyrimo išvadas.	Probleminis dėstymas, literatūros analizė, laboratorinių darbų atlikimas, diskusijos, projektinė užduotis.	
Gebės taikyti natūralios kalbos apdorojimo metodus, kelti hipotezes ir atlikti statistinę gautų duomenų analizę.		

Temos	Kontaktinio darbo valandos							Savarankiškų studijų laikas ir užduotys	
	Paskaitos	Konsultacijos	Seminarai	Pratybos	Laboratoriniai darbai	Praktika	Visas kontaktinis darbas	Savarankiškas darbas	Užduotys
1. Įvadas - Natūralios kalbos apdorojimo apibrėžimas - Kurso tikslai ir pagrindinės sąvokos - Trumpa istorija - Šiuolaikinių naudojimo sričių apžvalga	4				4		8	8	

2. Morfologija ir baigtinės technikos - Teksto skaidymas į teksto vienetus, žodžių kamienų išskyrimas, pagrindinių žodžio formų nustatymas - Fleksinė ir derivacinė morfologija - Baigtiniai automatai - Baigtiniai keitikliai - Lietuvių kalbos morfologinė analizė	4				4	8	8	Reguliarosios išraiškos Teksto skaidymas į teksto vienetus, žodžių kamienų išskyrimas, pagrindinių žodžio formų nustatymas
3. Žodžių nuspėjimas ir kalbos dalių žymėjimas - Žodžių skaičiavimas tekstuose - N-grama - Žodžių nuspėjimas - Taisyklėmis paremtas kalbos dalių žymėjimas - Atsitiktinis kalbos dalių žymėjimas - Įvardytos esybės	4				4	8	8	Žodžių nuspėjimas ir kalbos dalių žymėjimas. Įvardytų esybių radimas ir vizualizavimas.
4. Sintaksė ir gramatinis nagrinėjimas - Generatyvinė gramatika - Priklausomybių gramatika - Leksinis ir tikimybinis gramatinis nagrinėjimas - Lietuvių kalbos sintaksinė analizė	4				4	8	8	Sakinio sintaksinių struktūrų nustatymas. Sakinių gramatinis nagrinėjimas naudojant priklausomybių bei generatyvinę gramatiką.
5. Semantika - Semantikos reikšmė - Semantinė analizė - Pastovūs leksiniai ryšiai: hiponimai, antonimai, polisemai, sinonimai, hononimai - WordNet duomenų bazė - Informacijos paieška ir žodžio reikšmės parinkimas atsižvelgiant į kontekstą - Sentimentų analizė	4				4	8	8	Vektorinių ryšių tarp žodžių nustatymas. Teksto sentimentų nustatymas. Sentimentų duomenų rinkinio sudarymas.
6. Kontrolinis darbas	2				2	4	4	
7. Naudojimo sritys ir metodai - Diskursas - Dialogai ir pokalbių agentai - Teksto generavimas - Mašininio vertimo sistemos - Klasterizavimas - Informacijos suradimas - Tekstų klasifikavimas - Temų modeliavimas - Teksto santraukos sudarymas	8				10	18	18	Nesudėtingų sistemų kūrimas (projektinė užduotis)
8. Projekto rezultatų pristatymas	2					2	2	
9. Pasiruošimas baigiamajam testui							5	Literatūros kartojimas
Iš viso	32				32	64	69	

Vertinimo strategija	Svoris proc.	Atsiskaitymo laikas	Vertinimo kriterijai
Laboratoriniai darbai	30	Semestro metu (būtina atsiskaityti 50% laboratorinių darbų)	Laboratorinių darbų metu atliekamos trumpos užduotys. Galutiniam vertinimui naudojamas laboratorinių darbų įvertinimų aritmetinis vidurkis.
Kontrolinis darbas	20	Semestro metu	Kontrolinis darbas susideda iš 5 uždavinių.
Projektinė užduotis	20	Semestro metu	Projektinė užduotis atliekama dviejų-trijų studentų darbo grupėje (grupės dydis priklauso nuo projekto sudėtingumo). Darbo rezultatai pristatomi pranešimu paskaitos metu.
Egzaminas	30	Egzaminų sesijos metu (egzaminio laikymas eksternu nėra galimas)	Testas susideda iš 10 pasirenkamojo atsakymo klausimų.

Autorius	Leidimo metai	Pavadinimas	Periodinio leidinio Nr. ar leidinio tomas	Leidimo vieta ir leidykla ar internetinė nuoroda
Privaloma literatūra				
Lewis Tunstall, Leandro von Werra, Thomas Wolf	2022	Natural Language Processing with Transformers		O'Reilly Media, Inc.
Steven Bird, Ewan Klein, and Edward Loper	2009	Natural Language Processing with Python		O'Reilly Media

Papildoma literatūra				
Philipp Koehn	2020	Neural machine translation		Cambridge University Press.
Daniel Jurafsky and James Martin	2009	Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (second edition)		Prentice Hall
Aman Kedia, Mayank Rasu	2020	Hands-On Python Natural Language Processing: Explore tools and techniques to analyze and process text with a view to building real-world NLP applications		Packt Publishing Ltd
Bing Liu	2020	Sentiment analysis: mining opinions, sentiments, and emotions		Cambridge university press.
Nitin Indurkha, Fred J. Damerau.	2010	Handbook of natural language processing		Chapman and Hall/CRC
Mário Rodrigues, António Teixeira	2015	Advanced Applications of Natural Language Processing for Performing Information Extraction		Springer International Publishing
Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta, Harshit Surana	2020	Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems		O'Reilly Media, Inc
Karthiek Reddy Bokka, Shubhangi Hora, Tanuj Jain, Monicah Wambugu	2019	Deep Learning for Natural Language Processing: Solve your natural language processing problems with smart deep neural networks		Packt Publishing Ltd
Frank Millstein	2020	Natural Language Processing With Python: Natural Language Processing Using NLTK		Frank Millstein



COURSE UNIT (MODULE) DESCRIPTION

Course unit (module) title	Code
Natural Language Processing	

Lecturer(s)	Department(s) where the course unit (module) is delivered
Coordinator: Dr. Gražina Korvel Other(s): Dr. Gintautas Tamulevičius	Faculty of Mathematics and Informatics Institute of Data Science and Digital Technologies

Study cycle	Type of the course unit (module)
First	Optional

Mode of delivery	The period when the course unit (module) is delivered	Language(s) of instruction
Face-to-face	7 th semester	Lithuanian

Requirements for students	
Prerequisites: Math for Information Systems Engineering, Statistical Data Analysis Methods	Additional requirements (if any):

Course (module) volume in credits	Total student workload	Contact hours	Self-study hours
5	133	64	69

Purpose of the course unit (module): programme competencies to be developed

The aim of the course unit is to introduce the basic principles and theoretical issues of natural language processing. The course unit focuses on the computer use of natural language tools and techniques and their application in practice. It covers parsing, syntactic analysis, semantic interpretation, and language pragmatics. It also covers the main applications and techniques. Course laboratory activities highlight the main course topics. A group project investigates a single topic in greater depth.

Learning outcomes of the course unit (module)	Teaching and learning methods	Assessment methods
Ability to perform tasks on natural language processing independently and in a team.	Individual and group laboratory activities, group discussions, group projects, and presentations.	Laboratory activities, tests, group projects, and exam.
Ability to conduct proper literature on natural processing topics search and analysis, gather new knowledge, and apply it for problem solving.	Study of research literature, its analysis and evaluation, group project.	
Ability to plan, schedule, perform natural language processing experiments, evaluate the obtained results, and draw conclusions.	Lecture by teacher, study of research literature, laboratory activities, group discussion, group project.	
Ability to apply natural language processing algorithms, perform statistical analysis.		

Content: breakdown of the topics	Contact hours							Self-study work: time and assignments	
	Lectures	Tutorials	Seminars	Exercises	Laboratory work	Internship/work placement	Contact hours	Self-study hours	Assignments
1. Introduction <ul style="list-style-type: none"> - What is Natural Language Processing? - The tasks and fundamentals of the course - Brief history - Overview of Modern Applications 	4				4		8	8	Overview of current applications.

2. Morphology and Finite State techniques - Tokenization, stemming, lemmatization - Inflectional and derivational morphology - Finite-state automata - Finite-state transducers - The morphemic analysis of the Lithuanian language	4				4		8	8	Regular expressions. Tokenization, stemming, lemmatization.
3. Prediction and part-of-speech tagging - Counting words in Corpora - Simple N-grams - Word prediction - Rule-based part-of-speech tagging - Stochastic part-of-speech tagging - Named entities	4				4		8	8	Prediction and part-of-speech. Named entity recognition and visualization.
4. Syntax and parsing - Generative grammar - Parsing and generation with context-free grammars - Lexicalized and probabilistic parsing - Overview of Lithuanian Syntax	4				4		8	8	Find syntactic structures for sentences. Write a grammar parser and use it to find all the parses of a sentence.
5. Semantic - Representing meaning - Semantic analysis - Classical lexical relations: hyponymy, polysemy, synonymy, homonymy - WordNet: A Database of Lexical Relations - Word sense disambiguation and information retrieval - Sentiment analysis	4				4		8	8	Compute the similarity between the word vectors. Sentiment analysis of text.
6. The test	2				2		4	4	
7. Applications and techniques - Discourse - Dialogue and conversational agents - Generation - Statistical Alignment and Machine Translation - Clustering - Information retrieval - Text Categorization - Topic modeling - Keyword extraction	8				10		18	18	Design a simple dialog system (group project)
8. Presentation of project results	2						2	2	
9. Preparing for final test								5	Literature review
Total	32				32		64	69	

Assessment strategy	Weight, %	Deadline	Assessment criteria
Laboratory activities	30	Throughout the course (50% of laboratory work must be completed)	Tasks completed during laboratory work. The arithmetic mean of the grades of the laboratory assignments is used for the final grade.
Test	20	Throughout the course	The Test is composed of 5 open questions.
Group project	20	Throughout the course	The project is made in groups of 2-3 students (depending on task). The results of the project are presented during the lecture.
Exam	30	During examination session (external students are not eligible to take this exam)	The Test is composed of 10 multiple-choice questions.

Author	Year of publication	Title	Issue of a periodical or volume of a publication	Publishing place and house or weblink
Compulsory reading				
Lewis Tunstall, Leandro von Werra, Thomas Wolf	2022	Natural Language Processing with Transformers		O'Reilly Media, Inc.
Steven Bird, Ewan Klein, and Edward Loper	2009	Natural Language Processing with Python		O'Reilly Media

Optional reading				
Philipp Koehn	2020	Neural machine translation		Cambridge University Press.
Daniel Jurafsky and James Martin	2009	Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (second edition)		Prentice Hall
Aman Kedia, Mayank Rasu	2020	Hands-On Python Natural Language Processing: Explore tools and techniques to analyze and process text with a view to building real-world NLP applications		Packt Publishing Ltd
Bing Liu	2020	Sentiment analysis: mining opinions, sentiments, and emotions		Cambridge university press.
Nitin Indurkha, Fred J. Damerau.	2010	Handbook of natural language processing		Chapman and Hall/CRC
Mário Rodrigues, António Teixeira	2015	Advanced Applications of Natural Language Processing for Performing Information Extraction		Springer International Publishing
Sowmya Vajjala, Bodhisattwa Majumder, Anuj Gupta, Harshit Surana	2020	Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems		O'Reilly Media, Inc
Karthiek Reddy Bokka, Shubhangi Hora, Tanuj Jain, Monicah Wambugu	2019	Deep Learning for Natural Language Processing: Solve your natural language processing problems with smart deep neural networks		Packt Publishing Ltd
Frank Millstein	2020	Natural Language Processing With Python: Natural Language Processing Using NLTK		Frank Millstein