



STUDIJŲ DALYKO (MODULIO) APRAŠAS

Dalyko (modulio) pavadinimas	Kodas
Tekstų analizė naudojant <i>Python</i>	

Dėstytojas / a (-ai)	Padalinys (-iai)
Kuruojantis: doc. dr. Vytautas Rudžionis Kitas / a (-i): doc. dr. Kęstutis Driaunys, vyresn. mokslo darbuotoja dr. Egidija Kiškina, mokslo darbuotoja dr. Aušra Kairaitytė-Užupė	Kauno fakultetas Socialinių mokslų ir taikomosios informatikos institutas, Kalbų, literatūros ir vertimo institutas

Studijų pakopa	Dalyko (modulio) tipas
Pirmoji	Bendrųjų universitetinių studijų dalykas

Igyvendinimo forma	Vykdyimo laikotarpis	Vykdyimo kalba (-os)
Mišri (nuotolinė ir auditorinė)	Pavasario arba rudens semestras	Lietuvių

Reikalavimai studijuojančiajam	
Išankstiniai reikalavimai: Anglų kalba (bent B1)	Gretutiniai reikalavimai (jei yra): dalykas pirmiausia rekomenduojamas humanitarinių ir socialinių mokslų bakalauro programų studentams

Dalyko (modulio) apimtis kreditais	Visas studento darbo krūvis	Kontaktinio darbo valandos	Savarankiško darbo valandos
5	130	48	82

Dalyko (modulio) tikslas		
Studentai įgis įvadinį <i>Python</i> programavimo kalbos žinių ir natūralios kalbos apdorojimo bei teksto gavybos pagrindų žinių. Tai suteiks jiems galimybę automatizuoti skaitmeninės prigimties ir suskaitmenintų tekstinių duomenų surinkimą, analizuoti duomenis ir kelti naujus humanitarinių ir socialinių tyrimų klausimus.		
Dalyko (modulio) studijų rezultatai	Studijų metodai	Vertinimo metodai
Išmanys bazinius programavimo principus ir <i>Python</i> programinės kalbos pagrindus.	Paskaita (aiškinimas, pavyzdžių demonstravimas ir nagrinėjimas), literatūros ir šaltinių analizė, grupės diskusija, praktinis darbas.	Aktyvus dalyvavimas grupės diskusijose; pratybų užduoties atlikimas; egzaminas; darbas
Žinos bazines programavimo sąvokas (<i>duomenų struktūros, sąlygos sakiniai, ciklai, kintamieji ir funkcijos, kt.</i>) ir mokės jas taikyti sprendžiami uždaviniais.	Paskaita (aiškinimas, pavyzdžių demonstravimas ir nagrinėjimas), praktinis darbas, atvejų analizė.	Aktyvus dalyvavimas grupės diskusijose, atvejų analizėje; pratybų užduoties atlikimas; egzaminas; darbas
Mokės parašyti <i>Python</i> programą, kuri gali automatiškai išgauti, analizuoti ir vizualizuoti tekstyno informaciją.	Paskaita (aiškinimas, pavyzdžių demonstravimas ir nagrinėjimas), praktinis darbas, atvejų analizė.	Aktyvus dalyvavimas grupės diskusijose, atvejų analizėje; pratybų užduoties atlikimas; egzaminas; darbas
Mokės rinkti ir paruošti tekstinius duomenis ir susikurti tyrimui reikalingą tekstyną.	Paskaita (aiškinimas, pavyzdžių demonstravimas ir nagrinėjimas), praktinis darbas, atvejų analizė.	Aktyvus dalyvavimas grupės diskusijose, atvejų analizėje;

		pratybų užduoties atlikimas; egzaminas darbas
Gebės pristatyti savo tyrimo rezultatus raštu ir žodžiu naudodamas vizualines priemones.	Paskaita (aiškinimas, pavyzdžių demonstravimas ir nagrinėjimas), praktinis darbas, atvejų analizė.	Aktyvus dalyvavimas grupės diskusijose, atvejų analizėje; pratybų užduoties atlikimas; egzaminas darbas
Gebės identifikuoti tekstų analizės uždavinius, uždavinių sprendimo apimtis, laukiamų rezultatų tikslumo ribas, mokės įvertinti reikalingų uždavinio sprendimui išteklių poreikį.	Paskaita (aiškinimas, pavyzdžių demonstravimas ir nagrinėjimas), praktinis darbas, atvejų analizė.	Aktyvus dalyvavimas grupės diskusijose, atvejų analizėje; pratybų užduoties atlikimas; egzaminas darbas

Temos	Kontaktinio darbo valandos						Savarankiškų studijų laikas ir užduotys		
	Paskaitos	Konsultacijos	Seminarai	Pratybos	Laboratoriniai darbai	Praktika	Visas kontaktinis darbas	Savarankiškas darbas	Savarankiškai atliekamos užduotys
1. Programavimas. Python kalba. Kas yra programavimas. Kodėl reikalingos programavimo kalbos. Pagrindiniai programavimo kalbų principai ir elementai. Python programavimo priemonės. Dirbtinio intelekto platforma Anaconda, Python 3 ir <i>Natural Language Toolkit</i>	2			4			6	6	Užduotis: susipažinti su programavimo aplinka <i>Google Colab</i> . Python programinio kodo rašymas ir vykdymas. Pagrindiniai Python kalbos elementai
2. Natūralios kalbos apdorojimas (angl. Natural Language Processing/ NLP). Tekstas dirbtinio intelekto algoritmuose. Kas yra Natūralios kalbos apdorojimas (angl. NLP). Pagrindiniai NLP veikimo principai. Svarbiausi NLP algoritmai	2			2			4	4	Užduotis: susipažinimas su NLP grįstomis dirbtinio intelekto sistemomis <i>Google Cloud Natural Language AI; NLP Cloud, cedille.ai</i>
3. Tekstiniai ištekliai ir jų leksiniai ypatumai. Teksto gavimo būdai. Tekstyno sąvoka. Tekstynų apdorojimo principai ir paruošimas kompiuterinei analizei. Teksto gavimo būdai. Tekstynų suteikiamos galimybės.				2			2	4	Užduotis: sukurti teminį tekstyną panaudojant savo studijuojamai disciplinai būdingus šaltinius (pvz., interviu tekstai, dokumentai, filmų titrai, reklamų tekstai, straipsniai ir kt.)
4. Tekstyno sudarymas. Rankinis tekstų rinkimas ir apdorojimas. Automatinio teksto rinkimo būdai. Metaduomenų rinkimo būdai. Teisinio reguliavimo aspektai. Pirminis tekstyno sutvarkymas. Teksto vizualizavimas. Žodžių debesys.	2			6			8	14	Užduotis: automatiškai surinkti tekstus ir metaduomenis dominančia tema iš įvairių šaltinių (pvz., <i>Delfi; Scopis</i> duomenų bazės; atvirų duomenų archyvų ir kt.)
5. Baziniai tekstyno analizės būdai. Teksto skaidymas į leksinius ir morfologinius elementus. Teksto tokenizacija ir lemavimas. Tekstų turinio analizė ir klasifikavimas. Žodžių erdvė. <i>Žodžių krepšio</i> modelis. Semantiškai nereikšmingų žodžių pašalinimas. Tekstų klasifikavimas. Bajeso modeliai.	2			6			8	14	Užduotis: susikurto tekstyno paruošimas analizei. Tekstyno analizė tematikos požiūriu. Atliekamas automatinis tekstų skaidymas pasirinktos tematikos požiūriu.
6. Teksto klasterizavimas panašumų analizei. Teksto panašumo vertinimas. Tekstų panašumų matai. Atvejų pristatymai: automatizuota straipsnių atranka pagal jų turinį. Teksto autorystės identifikavimas.	4			6			10	16	Užduotis: savarankiškai atlikti tekstų klasterizavimo užduotį, atrenkant juos pagal turinį, autorių, laikotarpį arba kitus požymius (pvz., geografinę

									vieta, istorinius veikėjus, žinomus istorinius įvykius)
7. Pokalbių robotai/ asistentai? (Chatbotai). Dialogo modeliavimas. N-gramų modeliai. Žodžių sekų prognozavimo principai. Žodžių įterpimo idėjos. Tekstynų parengimas apmokymui naudoti neuroninius tinklus. BERT ir GPT tipo modeliai.	4			6				10	14
Konsultacija		2						2	
Egzaminas				2				2	6
Iš viso	16	2		34				52	78

Vertinimo strategija	Svoris proc.	Atsiskaitymo laikas	Vertinimo kriterijai
Projektinio darbo gynimas. Tekstyno sudarymas ir jo parengimas analizei	30 proc.	7–8 savaitė	Projektas skirtas patikrinti praktinio įgytų žinių pritaikymo gebėjimams. Vertinama 1–10 pažymių vertinimo skalėje: 10–9: Puikūs gebėjimai. Vertinimo lygmuo. 90–100 % teisingų atsakymų. 8–7: Geri gebėjimai, gali būti neesminių klaidų. Sintezės lygmuo. 70–89 % teisingų atsakymų. 6–5: Vidutiniai gebėjimai, yra klaidų. Analizės lygmuo. 50–69 % teisingų atsakymų. 4–3: Gebėjimai nesiekia vidutinių, yra (esminių) klaidų. Žinių taikymo lygmuo. 20–49 % teisingų atsakymų. 2–1: Netenkinami minimalūs reikalavimai. 0–19 % teisingų atsakymų.
2 kontroliniai darbai	30 proc.	15–16 savaitė	Vertinami studentų igūdžiai praktiškai taikyti įgytas žinias. Vertinama 1–10 pažymių vertinimo skalėje: 10–9: Puikūs gebėjimai. Vertinimo lygmuo. 90–100 % teisingų atsakymų. 8–7: Geri gebėjimai, gali būti neesminių klaidų. Sintezės lygmuo. 70–89 % teisingų atsakymų. 6–5: Vidutiniai gebėjimai, yra klaidų. Analizės lygmuo. 50–69 % teisingų atsakymų. 4–3: Gebėjimai nesiekia vidutinių, yra (esminių) klaidų. Žinių taikymo lygmuo. 20–49 % teisingų atsakymų. 2–1: Netenkinami minimalūs reikalavimai. 0–19 % teisingų atsakymų.
Egzaminas	40 proc.	Sesijos metu	Egzaminas skirtas patikrinti kurso metu įgytų žinių įsisavinimui. Vertinama 1–10 pažymių vertinimo skalėje: 10–9: Puikios žinios. Vertinimo lygmuo. 90–100 % teisingų atsakymų. 8–7: Geros žinios, gali būti neesminių klaidų. Sintezės lygmuo. 70–89 % teisingų atsakymų. 6–5: Vidutinės žinios, yra klaidų. Analizės lygmuo. 50–69 % teisingų atsakymų. 4–3: Žinios nesiekia vidutinių, yra (esminių) klaidų. Žinių taikymo lygmuo. 20–49 % teisingų atsakymų. 2–1: Netenkinami minimalūs reikalavimai. 0–19 % teisingų atsakymų.

Autorius (-iai)	Leidimo metai	Pavadinimas	Periodinio leidinio Nr. ar leidinio tomas	Leidykla ar internetinė nuoroda, leidinys yra atviras mokymosi išteklius (TAIP / NE)
Privaloma literatūra				
Benjamin Bengfort, Rebecca Bilbro, Tony Ojeda.	2018	Applied Text Analysis with Python		O'Reilly Media (TAIP)

Gabe Ignatow, Rada Mihalcea.	2017	An Introduction to Text Mining		SAGE Publications, Inc
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 https://github.com/PacktPublishing/Hands-On-Python-Natural-Language-Processing
Daiva Šveikauskienė	2022	Lietuvių kalbos gramatikos kompiuterizavimas		LKI, Vilnius
Jolanta Kovalevskatė, Erika Rimkutė, Jūratė Ruzaitė	2023	Tekstiniai ir jų išvestiniai produktai lietuvių kalbos mokymui(si) bei tyrimams: mokomoji priemonė		VDU, Kaunas
Papildoma literatūra				
Steven Bird, Ewan Klein, and Edward Loper	2009	Natural Language Processing with Python		O'Reilly Media https://www.nltk.org/book/
Yuli Vasiliev	2020	Natural Language Processing with Python and spaCy: A Practical Introduction		No Starch Press
Dipanjan Sarkar	2016	Text Analytics with Python – A Practical Real-World Approach to Gaining Actionable Insights from Your Data		Apress