

Studiju kursa nosaukums	Valodu tehnoloģijas
Studiju kursa kods	DatZW001
Zinātnes nozare	Datorzinātne un informātika
Zinātnes apakšnozare	Citas datorzinātnes un informātikas apakšnozares
Kredītpunkti(ECTS)	15
Kopējais kontaktstundu skaits	150
Lekciju stundu skaits	38
Semināru un praktisko darbu stundu skaits	112
Laboratorijas darbu stundu skaits	0
Studenta patstāvīgā darba stundu skaits	225
Kursa īstenošanas valoda	Latviešu
Kursa apstiprinājuma datums	04.04.2025
Atbildīgā struktūrvienība	Eksakto zinātņu un tehnoloģiju fakultāte

Kursa izstrādātājs(-i)

Dr.sc.comp., prof. Zane Bičevska

Priekšzināšanas

Kursa apguvei nepieciešamās priekšzināšanas atbilst studiju programmas uzņemšanas nosacījumiem un vispārējām zināšanām, prasmēm un kompetencēm, kas apgūtas iepriekšējā izglītības līmenī.

Studiju kursa anotācija

Papildu moduļa mērķis ir studentiem sniegt zināšanas par mūsdienu valodu tehnoloģijām. Valodu tehnoloģiju zināšanas var sniegt studentiem konkurences priekšrocības darba tirgū, jo uzņēmumi arvien biežāk meklē speciālistus, kas spēj izmantot un pielāgot mūsdienīgas valodu tehnoloģijas un mākslīgā intelekta risinājumus. Piemēram, mašīntulkošana, runas atpazīšana, uzvedņu izveide, automatizēta satura analīze un uzlabotas meklēšanas sistēmas tiek plaši izmantotas daudzās nozarēs. Valodu tehnoloģiju apguve palīdz studentiem izprast, kā dažādas valodas tiek apstrādātas digitālajā vidē un kā vislabāk tehnoloģijas izmantot darbā. Studējot valodu tehnoloģijas, studenti tiek aicināti domāt inovatīvi un radoši, risinot problēmas, kas saistītas ar valodas un tehnoloģiju integrāciju un lietojumu. Tas veicina viņu spēju risināt praktiskas problēmas, piemēram, tulkošanas produktivitātes celšanu vai valodas modeļu pielāgošanu specifiskām nozarēm.

Programmā paredzēts iekļaut Datorzinātņu programmā jau esošus kursus, bet nodrošinot padziļinātas konsultācijas tiem studentiem, kuriem nav iepriekš bijusi saskare ar datorzinātņu studiju kursiem.

Kursi:

- 1.Valodu tehnoloģiju lietojumi (3 ECTS)
- 2.Python programmēšanas valoda (3 ECTS)
- 3.Specseminārs par valodu tehnoloģiju tēmu (3 ECTS)
- 4.Kursa darbs par valodu tehnoloģiju tēmu (6 ECTS)

Kursa "Valodu tehnoloģiju lietojumi" mērķis ir apgūt jaunākās valodu tehnoloģijas teksta un runas, arī multimodālu datu apstrādei un prast šīs tehnoloģijas izmantot praktisku lietojumu izstrādē. Kursā tiks apskatīti dažāda veida transformeru modeļi dažāda veida

uzdevumu risināšanai. Galvenā uzmanība tiks veltīta multilingvāliem modeļiem, to izmantošanas, novērtēšanas un pielāgošanas iespējām. Praktisko iemaņu apgūšanai tiks izmantoti atvērtā pirmkoda valodas modeļi un programmēšanas ietvari. Tiks apskatītas arī galvenās metodes un risinājumi ģeneratīvo valodas modeļu un zināšanu bāzu/grafu kombinētai izmantošanai.

Kursa uzdevumi ir:

1. Iepazīstināt studējošos ar modernajām tehnoloģijām valodas sapratnei un valodas ģenerēšanai, runas atpazīšanai un runas sintēzei.
2. Padziļināt studējošo zināšanas un izpratni par valodu tehnoloģiju nozīmi plaša spektra IKT un mākslīgā intelekta (MI) risinājumu izstrādē.
3. Sniegt teorētiskas zināšanas un praktiskas iemaņas, kas nepieciešamas esošu MI risinājumu izmantošanā un integrēšanā, kā arī un jaunu, inovatīvu risinājumu radīšanā.

Kursa "Python programmēšanas valoda" mērķis ir sniegt studentiem pamatzināšanas par Python programmēšanas valodu.

Kursa uzdevumi:

- apgūt pamatzināšanas un programmēšanas iemaņas programmēšanas valodā Python;
- pilnveidot kursa dalībnieku programmēšanas prasmes;
- iepazīties ar pieejamajām Python programmu pakotnēm un to lietošanu;
- apgūt galvenos principus problēmu risināšanai, izmantojot programmēšanu.

Specseminārs ir Datorikas nodaļas iedibināta individualizēto studiju forma - zinātniskais seminārs. Specsemināram nedēļā paredzētas 90 kontaktminūtes. Konkrēto saturu un nodarbības veidu parasti nosaka katra konkrētā semināra vadītājs pēc saviem ieskatiem; valodu tehnoloģiju moduļa gadījumā specseminārs tiks veltīts valodu tehnoloģiju tematikai.

Specsemināra mērķis ir iemācīt studentiem piedalīties diskusijās ar pasniedzējiem un citiem studentiem par aktualitātēm datorzinātņu nozares pētniecībā un praksē. Specsemināra uzdevumi ir studentiem kopā ar pasniedzēju patstāvīgi studēt literatūru, risināt uzdevumus, izklāstīt kolēģiem izlasīto, kā arī mācīties savas specialitātes zināšanu metodiski pareizu izklāstu.

Kursa darba izstrāde ir patstāvīgs studenta darbs kursa darba vadītāja vadībā, kas atbilstoši Datorikas nodaļas apstiprinātajiem metodiskajiem ir aprakstīts un aizstāvēts īpašā kursa darbu aizstāvēšanas sēdē. Kursa darba mērķis ir apgūt pētnieciskā darba, akadēmiskās rakstīšanas, kā arī rezultātu prezentēšanas pamatprasmes. Kursa darba pamatu uzdevumi ir izveidot datorprogrammu vai veikt citu ar valodu tehnoloģijām saistītu aktivitāti, to aprakstīt gan akadēmiska darba, gan prezentācijas formā, kā arī aizstāvēt sēdē.

Papildu modulis tiek īstenots latviešu valodā.

Kursa plāns Pilna laika: klātiene

1. Mazie valodas modeļi: kodētāji un kodētāji-dekodētāji 4L 4Pd
2. Lielie valodas modeļi (LLM): dekodētāji. 4L 4Pd
3. Metodes zināšanu izguves un zināšanu grafu vadītai LLM izmantošanai 4L 2Pd
4. Runas atpazīšanas un sintēzes modeļi 4L 4Pd
5. Kursa projekta tēmas prezentācija 2S
6. Ievads. Mainīgie un datu tipi. 2L

7. Programmu izpildes kontroles konstrukcijas. Saraksti. 2L 2Pd
8. Datu struktūras. Funkcijas. 2L
9. Moduļi un Python standarta bibliotēka. 2L 2Pd
10. Darbs ar datnēm un direktorijām. 2L
11. Klases un objekti. Objektorientētā programmēšana. 2L 2Pd
12. NumPy bibliotēka. 2L
13. Datu analīze ar Pandas bibliotēku. 2L 2Pd
14. Funkcionālā programmēšana, Python ģenerātori. 2L
15. Dekoratori, konteksta pārvaldnieki. 2L
16. Datu vizualizācija. 2L 2Pd
17. Iknedēļas specseminārs 32S
18. Kursa darba sagatavošana 54Pd
- Kopā 38L 34S 78Pd

Studējošo patstāvīgo darbu organizācijas un uzdevumu raksturojums

Studentu patstāvīgais darbs ietver ar studiju kursa "Valodu tehnoloģiju lietojumi" četriem tematiskajiem blokiem saistīto valodas tehnoloģiju un to praktisko lietojumu apguvi, četru nelielu praktisko darbu izstrādi (individuāli), viena miniprojekta izstrādi (grupā), saistītās literatūras studijas, gatavošanos semināram un eksāmenam.

Teorētisko zināšanu nostiprināšanai un praktisko iemaņu apgūšanai ir paredzēti 2 praktiskie mājas darbi par kursa "Python programmēšanas valoda" tēmu, kurus studenti veic nelielās grupās.

Studenta patstāvīgais darbs specseminārā pamatā saistīts ar literatūras studijām, kuras izvēle notiek sadarbībā ar semināra vadītāju.

Atkarībā no konkrētā specsemināra specifikas, studenta patstāvīgais darbs var būt saistīts arī ar kāda neliela pētnieciska uzdevuma izveidi, izmantojot specseminārā apgūto. Patstāvīgi jā sagatavo arī prezentācija par lieteratūrā izlasīto (vai pētnieciskā uzdevuma rezultātu, ja to nosaka specsemināra noteikumi).

Kursa darba izstrādes ietvaros studentam semestra laikā:

1. Jāstudē ar veicamo izpēti saistītā literatūra.
2. Jāveic pētniecisks darbs.
3. Veiktais pētnieciskais darbs jāapraksta akadēmiska darba formātā.

Studiju rezultāti

Valodu tehnoloģiju lietojumi:

Zināšanas:

1. Izprot, analizē un salīdzina valodas apstrādes metodes, transformeru modeļu veidus un modalitātes, risināmos valodas apstrādes uzdevumus, aktuālos pētniecības virzienus un pamatmetodes. (EM11, EM12, EM21)
2. Identificē valodas apstrādes praktiskās un pētnieciskās problēmas, veic izpēti un analīzi. (EM13, EM22)

Prasmes:

3. Patstāvīgi piedāvā, izstrādā un novērtē dažāda veida valodas apstrādes risinājumus, izmantojot un pielāgojot modernākās tehnoloģijas, analizējot potenciālos riskus. (EM23, EM31, EM33, EM43, EM61)

Kompetence:

4. Izmanto iegūtās zināšanas un prasmes valodas tehnoloģijās balstītu kompleksu, inovatīvu IKT un MI risinājumu projektēšanā un izstrādē. (EM32, EM34)

5. Publiski izsakās par valodas tehnoloģiju jomas problemātiku, iespējām, metodēm un risinājumiem. (EM51, EM64)

Python programmēšanas valoda:

Zināšanas:

6. Pārzina Python programmēšanas valodas pamatus (EB11, EB24);

Prasmes:

7. Prot praktiski pielietot Python programmēšanas valodu un tās bibliotēku moduļus (EB22, EB24, EB35);

8. Orientējas Python programmu pirmkodā (EB22, EB23);

Kompetences:

9. Novērtē, kur un kā pielietot kursā gūtās Python programmēšanas valodas zināšanas (EB33, EB35);

10. Pielieto iegūtās zināšanas, patstāvīgi izstrādājot informācijas sistēmas (EB33, EB35, EB61).

Specseminārs:

Zināšanas

11. Izskaidro jaunākās publikācijas un/vai jaunākās tehnoloģijas izvēlētajā semināra jomā.

12. Izskaidro profesionālās saziņas un pētnieciskā darba principus un pieejas.

Prasmes

13. Prezentē izlasīto zinātniskajā literatūrā.

14. Diskutē par aktualitātēm nozarē

Kompetence

15. Formulē viedokli par seminārā dzirdēto un literatūrā izlasīto (eb11, eb12, eb13, eb14, eb15, EB21, eb22, eb23, eb24, eb25, eb31, eb32, eb33, eb34, eb35, eb41, eb42, eb43, eb44, eb51, eb52, eb53, eb54, eb55, eb61, eb62, eb63, eb64, eb65).

Kursa darbs:

Zināšanas

16. Izprot akadēmiska darba veidošanas principus. (eb11, eb12, eb13, eb14, EB15, eb22, EB23)

Prasmes

17. Veic pētījumu datorzinātņu nozarē, noformē pētījuma rezultātus akadēmiska darba formā. (eb54, EB55, EB61, eb62)

Kompetence

18. Patstāvīgi veic zinātnisku pētījumu un to dokumentē. (eb11, eb12, eb13, eb14, EB15, eb22, EB23, eb24, eb54, EB55, EB61, eb62)

Prasības kredītpunktu iegūšanai

Valodu tehnoloģiju pamati:

Starppārbaudījumi:

1. Praktiskais darbs (1. tēma) – 10%

2. Praktiskais darbs (2. tēma) – 10%

3. Praktiskais darbs (3. tēma) – 10%

4. Praktiskais darbs (4. tēma) – 10%

5. Seminārs: kursa miniprojekta idejas prezentācija – 10%

Noslēguma pārbaudījums:

6. Kursa miniprojekta izstrāde – 40%

7. Mutiskais eksāmens – 10%

Noslēguma pārbaudījuma kārtošanas priekšnosacījums: nokārtoti visi starppārbaudījumi.

Python programmēšanas valodas:

Starppārbaudījumi:

1. semestra laikā sekmīgi izpildīti un iesniegti 2 praktiskie mājas uzdevumi (40%),

2. iknedēļas pārbaudes uzdevumi (20%),

Noslēguma pārbaudījums:

3. sesijas laikā sekmīgi nokārtots eksāmens (mutisks un rakstisks, 40%),

4. aizpildīta LUIS anketa ar kursa novērtējumu (0%).

Sekmīga gala atzīme tiek saņemta pie nosacījuma, ka ir sekmīgi nokārtots eksāmens, ir sekmīgi ieskaitīti visi praktiskie mājas uzdevumi un students ir piedalījies vismaz 3 no 5 praktisko darbu nodarbībām.

Specseminārs:

Starppārbaudījums: Jāpiedalās kopējā diskusijā vismaz 2/3 nodarbību (50%)

Gala pārbaudījums: Eksāmens - Jāuzstājas ar patstāvīgu stāstījumu par semināra tēmu. (50%)

Kursa darbs:

Noslēguma pārbaudījums:

1. Kursa darba aizstāvēšana – 90 %

Papilduzdevums vērtējuma 10 saņemšanai (I-iespēja):

2. Kursa darba saturs atspoguļots angļu valodā uzrakstītā rakstā vismaz 2 pilnu A4 lappušu apjomā vai 4 pilnu A5 lappušu apjomā (noformējot atbilstoši zinātniska raksta prasībām, piemēram, pēc šāda parauga <https://www.bjmc.lu.lv/en/for-authors/instructions-for-authors/>) – 10%

Jāaizpilda LUIS anketa ar kursa novērtējumu. (0%)

Studiju rezultātu vērtēšanas kritēriji

Studiju kursa apguve tā noslēgumā tiek vērtēta saskaņā ar Latvijas Republikas normatīvajiem aktiem, vadoties pēc šādiem kritērijiem: iegūto zināšanu apjoms un kvalitāte, iegūtās prasmes, iegūtā kompetence atbilstīgi plānotajiem studiju rezultātiem.

Pārbaudījumu veidi	Studiju rezultāti																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Četri praktiskie darbi	+	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2. Seminārs: kursa miniprojekta idejas prezentācija	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-
3. Kursa miniprojekta izstrāde	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-
4. Mutiskais eksāmens	+	+	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-
5. Praktiskie mājas uzdevumi	-	-	-	-	-	+	+	+	+	+	-	-	-	-	-	-	-	-
6. Iknedēļas pārbaudes uzdevumi	-	-	-	-	-	+	+	+	+	-	-	-	-	-	-	-	-	-

7. Eksāmens (mutisks un rakstisks)	-	-	-	-	-	+	+	+	+	+	-	-	-	-	-	-	-	-
8. Daļība kopējā diskusijā	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	-	-	-
9. Eksāmens - Patstāvīgs stāstījums par semināra tēmu	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	-	-	-
10. Kursa darba aizstāvēšana	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+

Obligāti izmantojamie informācijas avoti

1. Nugues, P. M. Python for Natural Language Processing: Programming with NumPy, scikit-learn, Keras, and PyTorch. Springer, 2024
2. Tunstall, L., von Werra, L., Wolf, T. Natural Language Processing with Transformers. O'Reilly, 2022
3. Jurafsky, D. and Martin, J. H. Speech and Language Processing, 3rd ed. (<https://web.stanford.edu/~jurafsky/slp3/>)
4. Al Sweigart. "Automate the boring stuff with Python: practical programming for total beginners", 2nd edition. No Starch Press, 2019. <https://automatetheboringstuff.com/2e/>
5. David M. Beazley. "Python Distilled". Pearson, 2021.
6. PRASĪBAS NOSLĒGUMA DARBU (bakalaura, maģistra darbu, diplomdarbu un kvalifikācijas darbu) IZSTRĀDĀŠANAI UN AIZSTĀVĒŠANAI LATVIJAS UNIVERSITĀTĒ

Papildus informācijas avoti

1. Hugging Face NLP Course: <https://huggingface.co/learn/nlp-course/>
2. Hugging Face Audio Course: <https://huggingface.co/learn/audio-course/>
3. Charles Russell Severance. "Python for Everybody Exploring Data Using Python 3". 2016.
4. Wes McKinney. "Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter", 3rd edition. O'Reilly Media, 2022.

Periodika un citi informācijas avoti

1. ACL Anthology: <https://aclanthology.org>
2. Google's Python Class: <https://developers.google.com/edu/python>
3. Planet Python: <https://planetpython.org/>

Kursa ietvaros nav pieļaujams plaģiātisms un citi akadēmiskā godīguma pārkāpumi lūdzu iepazīties ar LU Noteikumiem par akadēmisko godīgumu. Šī kursa ietvaros ģeneratīvā mākslīgā intelekta (MI) rīku izmantošana ir atļauta izņēmuma gadījumos, ja to rakstiski ir norādījis un atļāvis šī kursa docētājs. Visos citos gadījumos nav pieļaujama mākslīgā intelekta ģenerētu materiālu (teksta, attēlu, audio, video u.c.) iesniegšana patstāvīgajos un grupu uzdevumos, pārbaudes darbā, eksāmenā vai jebkādā citā pārbaudes darbā, šāda veida materiālu iesniegšana tiks uzskatīta par neatļautu palīgīdzekļu izmantošanu