

**RTU studiju kurss "Dabiskās valodas apstrāde"**

33000 Datorzinātnes, informācijas tehnoloģijas un enerģētikas fakultāte

**Vispārējā informācija**

|                                                     |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
|-----------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Kods                                                | DEF725                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| Nosaukums                                           | Dabiskās valodas apstrāde                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                               |
| Studiju kursa statuss programmā                     | Obligātais/Ierobežotās izvēles                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
| Atbildīgais mācītbspēks                             | Gints Jēkabsons - Doktors, Docents                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |
| Apjoms daļās un kredītpunktos                       | 1 daļa, 6.0 kredītpunkti                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
| Studiju kursa īstenošanas valodas                   | LV, EN                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
| Anotācija                                           | Dabiskās valodas apstrāde ir starpdisciplināra zinātnes nozare, kas pēta metodes un modeļus, kuri ļauj datoram apstrādāt, interpretēt un ģenerēt cilvēku (dabisko) valodu. Studiju kursā studējošie apgūst dabiskās valodas apstrādes metodes, to realizēšanu programmatūrā un to dažādos praktiskos pielietojumus. Tiek apgūtas arī prasmes iegūtos rezultātus atbilstoši novērtēt un interpretēt.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
| Mērķis un uzdevumi, izteikti kompetencēs un prasmēs | Studiju kursa mērķis ir sniegt studējošajiem teorētiskās zināšanas un praktiskās iemaņas dabiskās valodas apstrādē un tās praktiskajos pielietojumos.<br>Studiju kursa uzdevumi:<br>- veicināt studējošo izpratni par dabiskās valodas apstrādes pamatiem un tās risinājumiem uzdevumiem;<br>- attīstīt studējošo prasmes identificēt un izvēlēties konkrētu uzdevumu risināšanai piemērotas dabiskās valodas apstrādes metodes;<br>- attīstīt studējošo prasmes praktiski realizēt un lietot dabiskās valodas apstrādes metodes konkrētu uzdevumu risināšanai, izmantojot atbilstošus programmatūras līdzekļus;<br>- attīstīt studējošo kompetenci pareizi novērtēt un interpretēt ar dabiskās valodas apstrādes metodēm iegūtos rezultātus.                                                                                                                                           |
| Patstāvīgais darbs, tā organizācija un uzdevumi     | Patstāvīgais darbs ietver individuālu teorētiskas un praktiskas ievirzes uzdevumu risināšanu, pielietojot studiju kursā iegūtās zināšanas par dabiskās valodas apstrādi, tās metodēm un atbilstošiem programmatūras līdzekļiem.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                         |
| Literatūra                                          | Obligātā/Obligatory:<br>1. Dan Jurafsky, James H. Martin. Speech and Language Processing 3rd ed. draft Prentice Hall, 2024, 1032 p. ( <a href="http://web.stanford.edu/~jurafsky/slp3/">http://web.stanford.edu/~jurafsky/slp3/</a> )<br>2. Steven Bird, Ewan Klein, Edward Loper. Natural Language Processing with Python ( <a href="http://www.nltk.org/book/">http://www.nltk.org/book/</a> )<br><br>Papildu/Additional:<br>1. Christopher Manning, Prabhakar Raghavan, Hinrich Schütze, Introduction to Information Retrieval, Cambridge University Press. 2008, 569 p..<br>2. Christopher Manning, Hinrich Schütze. Foundations of Statistical Natural Language Processing MIT Press, Cambridge, MA, 1999, 655 p..<br>3. Alexander Clark, Chris Fox, Shalom Lappin (eds.). The Handbook of Computational Linguistics and Natural Language Processing Wiley-Blackwell, 2010, 802 p. |
| Nepieciešamās priekšzināšanas                       | Programmēšana valodā Python.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                            |

**Studiju kursa saturs**

| Saturs                                                     | Pilna un nepilna laika klātienes studijas |                | Nepilna laika neklātienes studijas |                |
|------------------------------------------------------------|-------------------------------------------|----------------|------------------------------------|----------------|
|                                                            | Kontakt stundas                           | Patstāv. darbs | Kontakt stundas                    | Patstāv. darbs |
| Dabiskās valodas apstrādes pamati.                         | 6                                         | 6              | 0                                  | 0              |
| Teksta priekšapstrāde.                                     | 6                                         | 9              | 0                                  | 0              |
| Darbs ar teksta korpusiem.                                 | 6                                         | 10             | 0                                  | 0              |
| Vektoru semantika un jēdzientelpas.                        | 8                                         | 12             | 0                                  | 0              |
| Teksta dokumentu salīdzināšanas metodes.                   | 6                                         | 10             | 0                                  | 0              |
| Teksta dokumentu meklēšanas metodes.                       | 8                                         | 12             | 0                                  | 0              |
| Teksta klasificēšana.                                      | 10                                        | 15             | 0                                  | 0              |
| Valodas modeļi.                                            | 8                                         | 12             | 0                                  | 0              |
| Dažādi dabiskās valodas apstrādes praktiskie pielietojumi. | 6                                         | 10             | 0                                  | 0              |
| Kopā:                                                      | 64                                        | 96             | 0                                  | 0              |

**Sasniedzamie studiju rezultāti un to vērtēšana**

| Sasniedzamie studiju rezultāti              | Rezultātu vērtēšanas metodes                                                                                                                          |
|---------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------|
| Pārzina dabiskās valodas apstrādes pamatus. | Individuālie uzdevumi, eksāmens. Kritēriji: students demonstrē izpratni par jomas problēmu veidiem un to atbilstošajiem konceptuālajiem risinājumiem. |

|                                                                                                                                               |                                                                                                                                                                                                                     |
|-----------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Spēj identificēt un izvēlēties piemērotas metodes konkrētu problēmu risināšanai dabiskās valodas apstrādē.                                    | Individuālie uzdevumi, eksāmens. Kritēriji: students demonstrē kompetenci un prasmes risināt jomas tipiskas problēmas ar atbilstošām metodēm.                                                                       |
| Prot realizēt un pielietot dabiskās valodas apstrādes metodes konkrētu problēmu risināšanai, izmantojot atbilstošus programmatūras līdzekļus. | Individuālie uzdevumi, eksāmens. Kritēriji: students demonstrē izpratni par jomas tipisko metožu būtību un lietošanu, kā arī demonstrē spējas izstrādāt un izmantot risināmajai problēmai atbilstošu programmatūru. |
| Prot novērtēt un interpretēt ar dabiskās valodas apstrādes metodēm iegūtos rezultātus.                                                        | Individuālie uzdevumi, eksāmens. Kritēriji: students demonstrē kompetenci iegūto rezultātu interpretēšanā un tālākā apstrādē.                                                                                       |

#### ***Studiju rezultātu vērtēšanas kritēriji***

| Kritērijs             | % no kopējā vērtējuma |
|-----------------------|-----------------------|
| Individuālie uzdevumi | 50                    |
| Eksāmens              | 50                    |
| Kopā:                 | 100                   |

#### ***Studiju kursa plānojums***

| Daļa | KP  | Stundas  |          |         | Pārbaudījumi |        |       |
|------|-----|----------|----------|---------|--------------|--------|-------|
|      |     | Lekcijas | Prakt d. | Laborat | Ieskaite     | Eksām. | Darbs |
| 1.   | 6.0 | 20.0     | 2.0      | 0.0     |              | *      |       |