

**RTU studiju kurss "Specializētās datu apstrādes tehnoloģijas"**

33000 Datorzinātnes, informācijas tehnoloģijas un enerģētikas fakultāte

**Vispārējā informācija**

Kods	DE0756
Nosaukums	Specializētās datu apstrādes tehnoloģijas
Studiju kursa statuss programmā	Obligātais/Ierobežotās izvēles
Atbildīgais mācībspēks	Marina Uhanova - Doktors, Asociētais profesors
Mācībspēks	Aleksejs Jurenoks - Doktors, Asociētais profesors Padmaraj Nidagundi - Doktors, Docents
Apjoms daļās un kredītpunktos	1 daļa, 6.0 kredītpunkti
Studiju kursa īstenošanas valodas	LV
Anotācija	Studiju kursā tiek aplūkotas informācijas izguves sistēmu (jeb meklētājprogrammu) uzbūves principi, kā arī problēmas, kas saistītas ar lielu dokumentu kolekciju apstrādi. Studiju kursa ietvaros tiek izskatītas tēmas, kas saistītas ar tādiem informācijas izguves modeļiem kā vektoru telpas modelis, varbūtību modelis un valodas modelis, kā arī tiek aplūkoti kvantitatīvi meklētājprogrammu novērtēšanas kritēriji. Tiek izskatītas meklēšanas veikspējas uzlabošanai paredzētas indeksu struktūras, kas var būt izmantotas ne tikai informācijas izguves sistēmās, bet arī datubāžu sistēmu izstrādē.
Mērķis un uzdevumi, izteikti kompetencēs un prasmēs	Studiju kursa mērķis ir iepazīstināt studentus ar principiem un metodēm, kas tiek izmantotas informācijas izguves sistēmās. Studiju kursa uzdevumi ir iemācīt studentus implementēt efektīvas indeksu struktūras, kas paredzētas meklēšanas veikspējas uzlabošanai gan strukturētās, gan nestructurētās datu kolekcijās un kas var tikt izmantotas ne tikai informācijas izguves sistēmās, bet arī datubāžu sistēmu izstrādei.
Patstāvīgais darbs, tā organizācija un uzdevumi	Patstāvīgais darbs sastāv no metodisko materiālu un literatūras apgūšanas, kā arī laboratorijas darbu izpildes. Tiek piedāvāti astoņi laboratorijas darbu uzdevumi. Darbus ir jāizpilda pa posmiem saskaņā ar darba grafiku. Studenti patstāvīgi izstrādā šādus darbus: - invertētā indeksa izveide un izmantošana informācijas izguves sistēmās; - Būla vaicājumu apstrāde un izpilde informācijas izguves sistēmās; - TF-IDF svaru noteikšana un dokumentu ranžēšana; - K-grammu un permutācijas indeksu izmantošana aizstājējvaicājumu izpildei; - uz paplašināmās un lineāras jaucējtabulas balsītīte indeksi; - daudzdimensiju indeksu veidošana, izmantojot režģa failus un sadalītās jaucējfunkcijas; - daudzdimensiju indeksu veidošana, izmantojot kokveida struktūras; - punktu indeksu izveide daudzdimensiju datu apstrādei.
Literatūra	Obligātā/Obligatory: 1. Edward Sciore. Database Design and Implementation. Second Edition (Data-Centric Systems and Applications) 1st ed. Springer, 2020. 2. M. Tamer Özsu, Patrick Valduriez. Principles of Distributed Database Systems 4th ed. Springer, 2020. 3. ChengXiang Zhai, Sean Massung. Text Data Management and Analysis. A Practical Introduction to Information Retrieval and Text Mining. ACM Books, 2016. 4. C. Manning, P. Raghavan, and H. Schütze. Introduction to Information Retrieval. Cambridge University Press, 2008.  Papildu/Additional: 1. Abraham Silberschatz, Henry F. Korth, S. Sudarshan. Database System Concepts (Seventh Edition), McGraw-Hill, 2019. 2. Jeffrey D. Ullman Hector Garcia-Molina, Jennifer Widom. Database Systems The Complete Book. Pearson, 2014 3. Liema Wiese. Advanced Data Management. Berlin; Boston: De Gruyter, Oldenbourg, 2015. 4. Kleppmann, Martin. Designing data-intensive applications. O'Reilly Media, 2017. 5. Tanenbaum, Andrew S. Distributed systems : principles and paradigms. Pearson Prentice Hall, 2014.
Nepieciešamās priekšzināšanas	Kopu un varbūtību teorija, programmēšanas valodas, datu struktūras.

**Studiju kursa saturs**

Saturs	Pilna un nepilna laika klātienē studijas		Nepilna laika neklātienē studijas	
	Kontakt stundas	Patstāv. darbs	Kontakt stundas	Patstāv. darbs
Strukturēti un nestructurēti dati. Lielo datu veidi un apstrādes principi.	4	6	0	0
Informācijas izguve, tās pielietojumi un īpatnības.	2	1	0	0
Invertēta indeksa izveide.	6	9	0	0
Vārdnīcas un tolerantā informācijas izguve.	4	8	0	0
Vektoru telpas modelis un dokumentu ranžēšana.	6	10	0	0
Varbūtību un valodas modeļi.	6	10	0	0
Informācijas izguves sistēmu novērtēšana	6	6	0	0

Meklēšana tīmeklī. Saišu analīze. PageRank algoritms.	6	9	0	0
Datubāzu sistēmu implementācijas paņēmieni. Ierakstu glabāšana, fiksēta un mainīga garuma ieraksti.	6	9	0	0
Indeksu struktūras. B-koki, paplašināmās un lineāras jaucējtabulas.	6	10	0	0
Indeksu struktūras daudzdimensiju datiem. Režģa faili, sadalītas jaucējfunkcijas, kd-koki. R-koki un punktu indeksi.	6	9	0	0
Vaicājumu kompilēšana un izpildīšana.	6	9	0	0
<b>Kopā:</b>	<b>64</b>	<b>96</b>	<b>0</b>	<b>0</b>

#### **Sasniedzamie studiju rezultāti un to vērtēšana**

Sasniedzamie studiju rezultāti	Rezultātu vērtēšanas metodes
Spēj izskaidrot strukturēto un nestrukturēto datu glabāšanas un apstrādes principus.	Pozitīvs vērtējums par patstāvīgi sagatavotu referātu, sekmīgi nokārtoti kontroldarbi un eksāmens.
Spēj izskaidrot koncepcijas un algoritmus, kas tiek lietoti informācijas izguves sistēmās.	Sekmīgi nokārtoti kontroldarbi un eksāmens.
Spēj izstrādāt programmas, kas pielieto efektīvas indeksēšanas tehnikas nestrukturētu dokumentu kolekciju apstrādei.	Sekmīgi izpildīti un aizstāvēti laboratorijas darbu uzdevumi.
Spēj izstrādāt programmas, kas pielieto efektīvas indeksēšanas tehnikas strukturētu datu kolekciju apstrādei.	Sekmīgi izpildīti un aizstāvēti laboratorijas darbu uzdevumi.
Spēj izskaidrot meklētājprogrammu novērtēšanas kvantitatīvos kritērijus.	Sekmīgi nokārtoti kontroldarbi un eksāmens.

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

Kritērijs	% no kopējā vērtējuma
Laboratorijas darbu uzdevumi	40
Kontroldarbi	25
Referāts	10
Eksāmens	25
<b>Kopā:</b>	<b>100</b>

#### **Studiju kursa plānojums**

Daļa	KP	Stundas			Pārbaudījumi		
		Lekcijas	Prakt d.	Laborat	Ieskaite	Eksām.	Darbs
1.	6.0	32.0	0.0	32.0		*	