

Napredni modeli i baze podataka

Upute za 1. projekt Pretraživanje teksta u relacijskim bazama podataka akademska godina 2021./2022.

Projektni zadatak je podijeljen u dva dijela i oba se provode pomoću Edgara.

U prvom dijelu je, prilaganjem odgovora na postavljena pitanja, potrebno demonstrirati poznavanje teme. U drugom dijelu se u zadano vrijeme rješavaju zadaci nalik onima iz prvog dijela i odvijat će se kao on line provjera.

Rješenja oba dijela će se predavati u obliku teksta/slika putem sustava Edgar i ocjenjivati ručno od strane nastavnika

Studenti ne moraju pisati ni prilagati nikakvu dodatnu dokumentaciju.

Dio	Početak	Kraj	Kako pristupiti	
Prvi	13.10.2021.	26.10.2021. do 23:59.	https://edgar.fer.hr/ Public exams -> 1. project (text search) - part one	Provjere će biti dostupne za pokretanje kada nastupi trenutak Početak .
Drugi	27.10.2021. od 21:00 -22:00		https://edgar.fer.hr/ Public exams -> 1. project (text search) - part two	

1. Korištenje PostgreSQL-a

Kao SUBP u ovom projektu koristi se PostgreSQL. Prema uputama objavljenima u datoteci **Instalacija PostgreSQL i streamService baze - upute.pdf** u repozitoriju **Zimski semestar 2021./2022.** instalirajte PostgreSQL SUBP.

Nakon instalacije, SQL naredbe nad PostgreSQL sustavom možete obavljati pomoću nekog od SQL editora prilagođenih za rad s PostgreSQL-om npr. PgAdmin (koji ste vjerojatno instalirali zajedno s instalacijom SUBP).

Zbog nepostojanja cjelovite podrške za potpunu pretragu teksta za hrvatski jezik koristite tekstove na engleskom jeziku. Iznimka su dijelovi teksta u kojima je u svrhu demonstracije rješenja potrebno koristiti riječi s npr. palatalima.

Da biste mogli koristiti funkcije za Full Text Search i Fuzzy Text Search morate uključiti module **fuzzystrmatch** i **pg_trgm**. Module je potrebno uključiti u svakoj bazi podataka u kojoj ih namjeravate koristiti.

„Uključuju/registiraju“ se izvođenjem sljedećih naredbi:

```
CREATE EXTENSION fuzzystrmatch;      -- (soundex, levenshtein, metaphone)
CREATE EXTENSION pg_trgm;            -- (similarity , show_Trgm,..., %, <->)
```

2. Testni skup podataka

Da biste riješili zadatke trebat će vam tekstualni podaci, pa ćete jedan dio vremena morati izdvojiti za pribavljanje, obradu i punjenje baze podataka podacima. Prije nego li krenete s tim, proučite zadatke i dobro isplanirajte shemu. Npr. za neke od zadataka bi tekstualni podaci morali imati nekoliko tekstualnih atributa jer inače nećete moći pokazati ono što se zadatkom traži.

Detaljnije upute kao i linkove na kojima ćete pronaći podatke pročitajte u prvom zadatku u Edgaru.

3. Teme

U nastavku su linkovi na Postgresovu dokumentaciju koju, uz predavanja, treba proučiti prije rješavanja zadataka.

Tema: konfiguriranje parametara pretrage teksta, parser, rječnici

<https://www.postgresql.org/docs/current/sql-createtsconfig.html>
<https://www.postgresql.org/docs/current/sql-altertsconfig.html>
<https://www.postgresql.org/docs/current/textsearch-parsers.html>
<https://www.postgresql.org/docs/current/sql-createtsdictionary.html>
<https://www.postgresql.org/docs/current/sql-altertsdictionary.html>

Tema: podatkovni tipovi za prezentaciju dokumenata u normaliziranom obliku, operatori i funkcije za FTS

<https://www.postgresql.org/docs/current/textsearch-features.html>
<https://www.postgresql.org/docs/current/functions-textsearch.html>

Tema: rangiranje dokumenata

<https://www.postgresql.org/docs/current/textsearch-controls.html#TEXTSEARCH-RANKING>

Tema: približna pretraga teksta (Fuzzy text search)

<https://www.postgresql.org/docs/current/fuzzystrmatch.html>
<https://www.postgresql.org/docs/current/pgtrgm.html#id-1.11.7.40.5>