**1.projekt iz kolegija Napredni modeli i baze podataka**

Za prvi projekt je izrađena web aplikacija koja predstavlja jednostavnu tražilicu dokumenata. Kao predmet pretraživanja korišteni su nazivi filmova. Kako bi se moglo ispitati postojanje određenih filmova, trebalo je omogućiti više funkcionalnosti tražilice.

Ponajprije je trebalo omogućiti unos filmova u bazu. Pri tome se unosi naziv filma, redatelj, radnja, žanr, trajanje, boja te glavni glumac. Tek unosom filmova u bazu moguće je obaviti pretraživanje u bazi pohranjenog sadržaja koji je nužno morao biti na engleskom jeziku.

Ostvarene su tri vrste pretrage:

* ***Exact string matching*** – dohvat dokumenata koji sadrže traženi uzorak u neizmjenjenom obliku,
* ***Use dictionaries*** - dohvat dokumenata koji sadrže traženi uzorak u normaliziranom obliku,
* ***Fuzzy string matching*** - dohvat dokumenata koji sadrže približno jednak uzorak

Uzorke se može povezati logičkim operatorom AND ili OR. Različite vrste pretrage su dale različite rezultate koji su vidljivi i kod uporabe operatora.

Još jedna funkcionalnost tražilice jest i analiza upita. Pri tome je moguće izabrati želi li se analiza po danima ili satima za odabrani vremenski period. Ovaj dio vježbe je fokusiran na ostvarenje funkcionalnosti pivotiranjem.

U sklopu vježbe stvena je shema movies u kojoj su dvije tablice: filmovi i upiti. Prva je stvorena naredbom CREATE TABLE:

CREATE TABLE movies.filmovi

(

id serial NOT NULL,

naziv character varying(40) NOT NULL,

redatelj character varying(40) NOT NULL,

radnja character varying(200),

zanr character(20),

vrijeme\_trajanja smallint,

datum\_unosa date NOT NULL,

vrijeme\_unosa time without time zone NOT NULL,

u\_boji boolean,

glavni\_glumac character varying(50),

naziv\_ts tsvector,

CONSTRAINT filmovi\_pkey PRIMARY KEY (id)

),

dok je druga stvorena korištenjem:

CREATE TABLE movies.upiti

(

id\_upit serial NOT NULL,

upit character(100) NOT NULL,

datum\_unosa date,

vrijeme\_unosa timestamp without time zone,

CONSTRAINT upiti\_pkey PRIMARY KEY (id\_upit)

)

U prvoj tablici je bitno naglasiti važnost atributa naziv\_ts tsvector koji je TSVector tipa, a korišten je radi ubrzavanja pretraživanja. Naziv filma se normalizira i sprema u tablicu kako bi se prilikom pretraživanja isti mogao samo dohvatiti, a ne svaki put ponovno izračunavati.

Druga tablica je korištena za spremanje upita za analizu radi jednostavnijeg dohvata traženih ntorki.

Kreirana su dva indeksa radi ubrzavanja pretraživanja. Jedan je kreiran korištenjem GIN indeksa koji je puno brži nego GIST za pretragu i kreiran je nad atributom naziv\_ts. Drugi je pak kreiran korištenjem GIST indeksa koji zna brzo upariti vektor trigrama dokumenata s vektorom znakovnog niza koji se traži.

Skripta index.php je praktički glavni dio u kojem je ostvareno pretraživanje uz parsiranje unesenog znakovnog niza. U skripti dodavanje.php je ostvareno dodavanje zapisa, a skripta analiza.php se brine oko analize obavljenih upita. Ostatak je uglavnom orijentiran na dizajn aplikacije i pripadne css i html datoteke.