

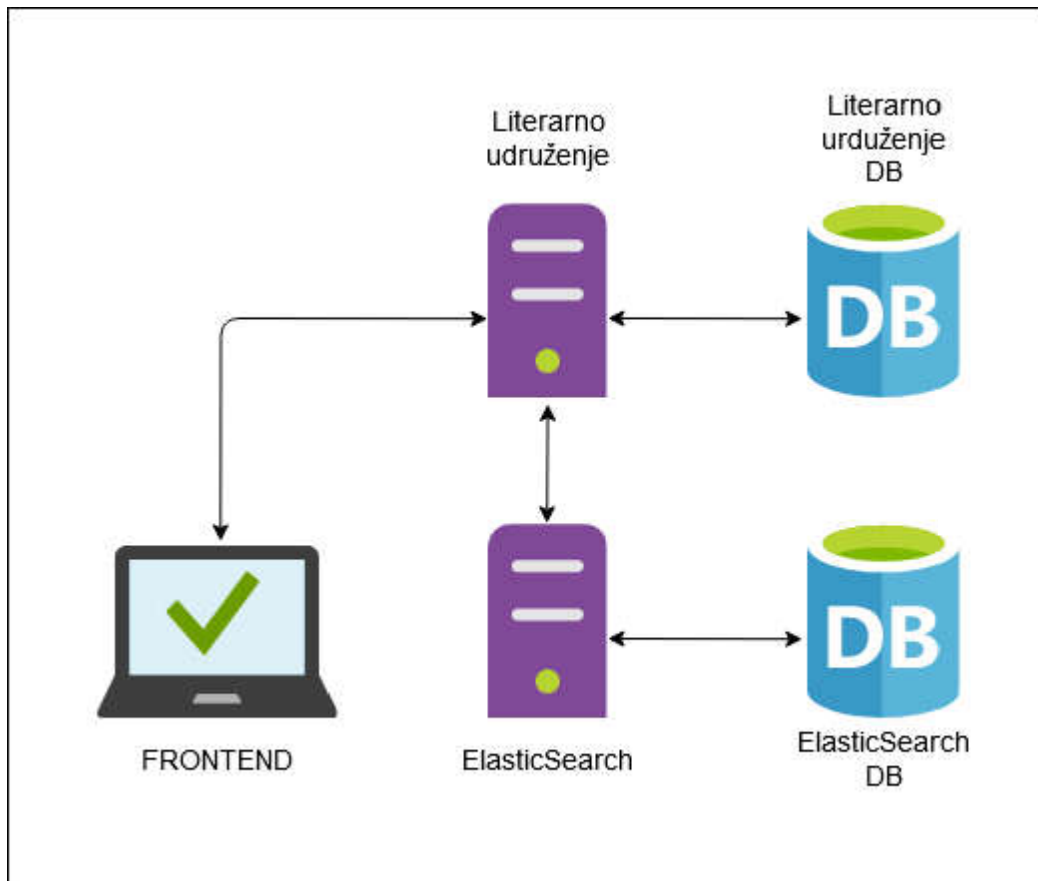
Upravljanje digitalnim dokumentima

Projekat iz predmeta **Upravljanje digitalnim dokumentima** predstavlja izradu aplikacije koja će omogućiti pretraživanje knjiga napisanih na srpskom jeziku, u okviru **Literarnog udruženja**. Literarno udruženje je Spring Boot aplikacija, koja omogućava lektorisanje, revizije, uređivanje i izdavanje knjiga. Kako bi se obezbedilo pretraživanje knjiga po raznim kriterijumima, koristi se Elasticsearch, koji omogućava pretraživanje knjiga pomoću upita.

Arhitektura rešenja

Rešenje se sastoji iz više delova:

- Frontend – React aplikacija koja predstavlja korisnicki interfejs, pomoću koje se zadaju upiti i prikazuju rezultati upita
- Literarno udruženje – Spring Boot aplikacija koja obavlja celokupnu logiku i koja prosleđuje zahteve Elasticsearch-u.
- Elasticsearch – Platforma koja omogućava pretraživanje knjiga na osnovu zadatih upita



Literarno udruženje komunicira sa frontendom, kao i sa Elasticsearch-om putem REST API-ja. Literarno udruženje će slušati na portu 8080, frontend aplikacija na portu 3000, dok će Elasticsearch slušati na portu 9200.

ElasticSearch

Izvršavanjem komande **elasticsearch.bat**, koja se nalazi u **bin** folderu Elasticsearch-a, se pokreće Elasticsearch. Komandom **curl -X GET http://localhost:9200/** moguće je proveriti da li je Elasticsearch pokrenut. Elasticsearch će koristiti analyzer za srpski jezik, ElasticsearchTemplate I ElasticsearchRepository. Neophodno je obavestiti Elasticsearch da koristi i srpski analyzer, pored podrazumevanog engleskog. Na slici ispod je prikazano kako obavestiti Elasticsearch da koristi i srpski analyzer.

```
1  {
2    "mappings": {
3      "properties": {
4        "content": {
5          "type": "text",
6          "fields": {
7            "sr": {
8              "type": "text",
9              "analyzer": "serbian"
10             },
11             "en": {
12               "type": "text",
13               "analyzer": "english"
14             }
15           }
16         }
17       }
18     }
19   }
20 }
```

Plagiator

Za detekciju potencijalnih plagijarizama će se koristiti alat dostupan na <https://github.com/chenejac/plagiator>. Komunikacija će se obavljati preko REST API-ja.

Indexing unit

- Prilikom kreiranja nove knjige, indexing unit će izgledati kao na slici ispod.

```
1  {
2    "title": "title text",
3    "authors": [
4      {
5        "firstName": "John",
6        "lastName": "Doe"
7      },
8      {
9        "firstName": "John",
10       "lastName": "Doe"
11     }
12   ],
13   "genre": "romance",
14   "isbn": 1234,
15   "place": "Narnia",
16   "year": 1453,
17   "pages": 325,
18   "synopsis": "i hate writing",
19   "keyWords": "love kiss",
20   "publisher": "Book Factory"
21 }
22
23
```

- Dodavanje beta čitalaca će izgledati kao na slici ispod (šalje se id čitaoca, kao i njegove koordinate).

```
1  {
2    "userId": 123,
3    "latitude": 41.01,
4    "longitude": 28.58
5  }
6
7
8
```

Geoprostorna pretraga

ElasticSearch ima podršku za prostornu pretragu, te će se koristiti za pronalazak beta čitalaca koji su dovoljno udaljeni od autora knjige. Koristiće se `geo_point`, koji će predstavljati lokacije čitalaca. ElasticSearch će za svakog čitaoca čuvati njegov id i koordinate na kojima se nalazi. Prilikom filtriranja kandidata za beta čitaoca, poslaće se zahtev Literarnom udruženju da se dobave svi autori knjige, kao i njihove koordinate. Nakon toga će se izvršiti pretraga i biće pronađeni odgovarajući čitaoci. Na slici ispod je prikazan primer geoprostorne pretrage.

```
1  {
2    "query":
3    {
4      "bool" :
5      {
6        "must":
7        {
8          "match_all": {}
9        },
10       "filter": {
11         "distance": {
12           "value": 200,
13           "latitude": 41.01,
14           "longitude": 28.58
15         }
16       }
17     }
18   }
19 }
```

Serbian Analyzer

Pretprocesiranje knjiga na srpskom jeziku će se vršiti pomoću plugin-a koji se nalazi na linku <https://github.com/chenejac/udd06>. Neophodno je klonirati ovaj repozitorijum, zatim izvršiti komandu **gradlew clean build**, koja će kreirati *serbian-analyzer-1.0-SNAPSHOT.zip* unutar **build/distributions/** foldera. Nakon toga se treba pozicionirati u bin folder ElasticSearch-a i izvršiti komandu **./elasticsearch-plugin install file:<absolute path of distribution archive>**.

Skladištenje podataka

Baza podataka Literarnog udruženja će čuvati podatke o knjigama i čitaocima.

Svaka knjiga će sadržati:

- Naslov
- Jednog ili više autora
- Žanr
- ISBN
- Mesto izdavanja
- Godina izdavanja
- Broj stranica
- Sinopsis
- Ključni pojmovi
- Izdavač

Svaki čitalac ima:

- Ime
- Prezime
- Grad
- Državu
- Email
- Žanrovi za koje je zainteresovan
- Da li je beta čitalac

Željko Milošević
E2 131/2020