|  |  |  |
| --- | --- | --- |
|  | УНИВЕРЗИТЕТ У НОВОМ САДУ  **ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА У НОВОМ САДУ** |  |

Невена Глигоров

**DebugIt - студентска QA платформа која интегрише *Elasticsearch* механизам претраге**

ДИПЛОМСКИ РАД

- Основне академске студије -

Нови Сад, 2024.

Образац **Q2.НА.15**-**04** - Издање 1

**КЉУЧНА ДОКУМЕНТАЦИЈСКА ИНФОРМАЦИЈА**

|  |  |  |  |
| --- | --- | --- | --- |
| Редни број, **РБР**: | |  | |
| Идентификациони број, **ИБР**: | |  | |
| Тип документације, **ТД**: | | Монографска публикација | |
| Тип записа, **ТЗ**: | | Текстуални штампани документ | |
| Врста рада, **ВР**: | | Завршни-bachelor рад | |
| Аутор, **АУ**: | | Невена Глигоров | |
| Ментор, **МН**: | | доц. др Дуња Врбашки | |
| Наслов рада, **НР**: | |  | |
| Језик публикације, **ЈП**: | | Српски (ћирилица) | |
| Језик извода, **ЈИ**: | | Српски / Енглески | |
| Земља публиковања, **ЗП**: | | Србија | |
| Уже географско подручје, **УГП**: | | Војводина | |
| Година, **ГО**: | | 2024. | |
| Издавач, **ИЗ**: | | Ауторски репринт | |
| Место и адреса, **МА**: | | Факултет Техничких Наука (ФТН), Д. Обрадовића 6, 21000 Нови Сад | |
| Физички опис рада, **ФО**: (поглавља/страна/ цитата/табела/слика/графика/прилога) | | n/n/n/n/n/n/n | |
| Научна област, **НО**: | | Електротехничко и рачунарско инжењерство | |
| Научна дисциплина, **НД**: | | Примењене рачунарске науке и информатика | |
| Предметна одредница/Кqучне речи, **ПО**: | | Неколико кључних термина из рада, односно предмета изучавања на који се рад односи. | |
| **УДК** | |  | |
| Чува се, **ЧУ**: | | Библиотека ФТН, Д. Обрадовића 6, 21000 Нови Сад | |
| Важна напомена, **ВН**: | |  | |
| Извод, **ИЗ**: | | Кратак опис рада који стаје у три до шест реченица. Треба навести проблем који се решавао, како се проблем решавао и који су резултати. | |
| Датум прихватања теме, **ДП**: | |  | |
| Датум одбране, **ДО**: | |  | |
| Чланови комисије, **КО**: | Председник: |  |
|  | Члан: |  | Потпис ментора |
|  | Члан, ментор: | др Дуња Врбашки, доцент |  |

**KEY WORDS DOCUMENTATION**

|  |  |  |  |
| --- | --- | --- | --- |
| Accession number, **ANO**: | |  | |
| Identification number, **INO**: | |  | |
| Document type, **DT**: | | Monographic publication | |
| Type of record, **TR**: | | Textual material, printed | |
| Contents code, **CC**: | | Bachelor thesis | |
| Author, **AU**: | | Nevena Gligorov | |
| Mentor, **MN**: | | Dunja Vrbaški, PhD, assist. prof. | |
| Title, **TI**: | |  | |
| Language of text, **LT**: | | Serbian (cyrillic script)/Serbian (latin script) | |
| Language of abstract, **LA**: | | Serbian/English | |
| Country of publication, **CP**: | | Serbia | |
| Locality of publication, **LP**: | | Vojvodina | |
| Publication year, **PY**: | | 2024 | |
| Publisher, **PB**: | | Author reprint | |
| Publication place, **PP**: | | Faculty of Technical Sciences, D. Obradovića 6, 21000 Novi Sad | |
| Physical description, **PD**: (chapters/pages/ref./tables/pictures/graphs/appendixes) | | n/n/n/n/n/n/n | |
| Scientific field, **SF**: | | Electrical and computer engineering | |
| Scientific discipline, **SD**: | | Applied computer science and informatics | |
| Subject/Key words, **S**/**KW**: | |  | |
| **UC** | |  | |
| Holding data, **HD**: | | Library of the Faculty of Technical Sciences, D. Obradovića 6, 21000 Novi Sad | |
| Note, **N**: | |  | |
| Abstract, **AB**: | |  | |
| Accepted by the Scientific Board on, **ASB**: | |  | |
| Defended on, **DE**: | |  | |
| Defended Board, **DB**: | President: |  |
|  | Member: |  | Menthor's sign |
|  | Member, Mentor: | Dunja Vrbaški, PhD, assist. prof. |  |

|  |  |  |
| --- | --- | --- |
|  | УНИВЕРЗИТЕТ У НОВОМ САДУ ⚫ **ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА**  21000 НОВИ САД, Трг Доситеја Обрадовића 6 | Број: |
|  |
| **ЗАДАТАК ЗА ИЗРАДУ ДИПЛОМСКОГ (BACHELOR) РАДА** | Датум: |
|  |

Za sada ostaviti ovako, prazno, ali me podsetite pre finalne verzije da pošaljem ovu stranicu.

Spisak skraćenica

FTP - *File transfer protocol*

QA - *questions and answers*

Sadržaj

[1. Uvod 7](#_Toc175693236)

[1.1 Šta pisati u uvodu? 7](#_Toc175693237)

[2. Istorijat 8](#_Toc175693238)

[3. Značaj i primena 11](#_Toc175693239)

[3.1 Postojeća rešenja 12](#_Toc175693240)

[*Stack Exchange* i *Stack Overflow* 12](#_Toc175693241)

[*Discuss The Elastic Stack* 12](#_Toc175693242)

[*GitHub community* 12](#_Toc175693243)

[Ostalo 12](#_Toc175693244)

[4. *Elasticsearch* mehanizam 13](#_Toc175693245)

[4.1 Šta je *Elasticsearch* mehanizam i kako je nastao? 13](#_Toc175693246)

[4.2 Struktura i način rada *Elasticsearch* mehanizma 13](#_Toc175693247)

[5. Zaključak 15](#_Toc175693248)

[Literatura 16](#_Toc175693249)

[Dodatak A 17](#_Toc175693250)

[Dodatak B 18](#_Toc175693251)

[Podaci o kandidatu 19](#_Toc175693252)

# Uvod

U ovom dokumentu su navedene neke: sadržajne, tehničke i stilske smernice za pisanje završnog rada. Ujedno predstavlja i šablon za izradu i pisanje.

Ostale napomene se nalaze na adresi: [Zavrsni radovi - napomene](https://docs.google.com/document/d/e/2PACX-1vQ9_rgcnXK0Lw5_l6W16CbQ2MQudjqy9lDFMzwJmakopojXnTAzVDxl6xc-M0xCeUS3zr9VKOxlliRm/pub). Obavezno pogledati i ovaj dokument koji se s vremena na vreme dopunjava.

Možete korisiti ovaj dokument kao šablon, možete napraviti novi, a možete pisati i u LaTex-u. Rad se, na kraju, komisiji i biblioteci, predaje u pdf-u.

Prva strana, KDI tabele i zadatak rada treba da ostanu u datom formatu. Ostalo može biti formatirano drugačije.

Sam rad može biti napisan na ćirilici ili latinici s tim da naslovna strana i KDI moraju biti npisane kao što je navedeno.

## Šta pisati u uvodu?

Uvod sadrži kratak opis problema koji se rešava ili oblast koja se istražuje.

Može se navesti i motivacija za izbor određene teme, kontekst, značaj i povezanost sa drugim oblastima ili temama.

Na kraju se obično jednom rečenicom navodi šta koje poglavlje, u nastavku rada, opisuje.

Savet: Ne morate krenuti sa pisanjem odmah od Uvoda. To ume da blokira nekad. Slobodno krenite od glavne stvari pa ćete usput ili na kraju sami shvatiti šta je prirodno da piše u uvodu.

# Istorijat

Alan Emtage, administrator sistema za departman informacionih tehnologija i student kanadskog univerziteta McGill, čije je sedište u Montrealu, u septembru 1990. godine objavljuje prvi Internet mehanizam pretrage (*eng. search engine*) pod nazivom *Archie*. Ime *Archie* potiče od reči „*Archive*“, po uzoru na dugačak i mukotrpan proces pretrage kroz koji je Alan prolazio kako bi razvio svoj softver [1].

*Archie* korisnički interfejs je poprilično jednostavan, kao što je prikazano na slici 1, uzimajući u obzir da je njegova glavna namena pretraživanje fajlova na FTP sajtovima koji se nalaze na Internetu.



Slika 1. Archie interfejs [2]

Univerzitet je pred Alana postavio težak zadatak u to veme, a to je da kreira softver uz pomoć kog je moguće povezati se na Internet. Novokreirani *Archie* je predstavljao indeks FTP fajlova i njegovi korisnici su uz pomoć jednostavnih upita mogli da pretražuju fajlove. Ovi upiti su morali da budu dosta precizni, kako bi korisnici dobili ono što su želeli, pošto *Archie* nije imao podršku za upite napisane prirodnim jezikom (*eng. natural* language), niti je indeksirao sadržaj fajlova. Kako nije imao mogućnost da dâ korisnicima informaciju o tome šta se nalazi u fajlovima koji su im ponuđeni kao rezultat pretrage, korisnici su te fajlove morali da skinu (*eng. download*) na svoj računar i pogledaju njihov sadržaj.

Do 1993, *Archie* je postao veoma popularan i korišćen od strane drugih univerziteta širom sveta [3].

Nekoliko godina i mehanizama pretrage kasnije, apsolutnu dominaciju u Internet svetu ima Yahoo. Bilo da je ime *backronym* za „*Yet Another Hierarchically Organized Oracle*“ ili „*Yet Another Hierarchical Officius Oracle*“, njegova struktura je ista, takva da predstavlja direktorijum koji se sastoji od Internet sajtova, prilikom čega su sajtovi organizovani hijerarhijski. Primer interfejsa videti na slici 3 [4].



Slika 4, Yahoo! interfejs

*Yahoo* se temelji na ideji dvojice studenata sa Stanford Univerziteta, Jerry Yang i David Filo, čija je zamisao bila da naprave direktorijum pod nazivom „*Jerry and David's Guide to the World Wide Web*“, uz pomoć kojeg bi se navigirali do svojih omiljenih sajtova.

Zahvaljujući Internet pregledaču *Navigator* koji je postavio link do njihovog projekta na vrh svoje stranice, direktorijum koji su napravili Jerry i David polako postaje sve popularniji, tačnije, postaje toliko popularan da odlučuju da svoju ideju pretoče u biznis.

Iako je postojalo dosta drugih mehanizama pretrage, poput: *Lycos*, *AskJeeves*, *Excite*, *AOL*, *AltaVista*, *Infoseek* i drugih, *Yahoo* je stajao na samom vrhu i niko mu nije bio ni blizu. 1998. godine dvojica studenata odlaze u *Yahoo* sa ponudom da im prodaju svoj biznis za milion dolara, međutim *Yahoo* ih je odbio. Ono što tada *Yahoo* nije znao jeste da je odbi0 budućeg giganta koji i dan danas vlada Internetom, čiji primer interfejsa je moguće videti na slici 4 [6].



Slika 4. Google interfejs [8]

*Google* je polako postajao glavni protivnik *Yahoo* kompanije i *Yahoo*, imajući to u vidu, ponudio je *Google* da integriše njihov mehanizam pretrage u okviru svoje Internet stranice. Ovim potezom, *Google* je dobio promociju od strane kompanije *Yahoo*, kojoj to nije bila zamisao, i time se polako peo ka samom vrhu.

Larry Page i Sergey Brin su bili studenti Stanford Univerziteta kada su napravili prvi mehanizam pretrage koji je radio po principu analiziranja linkova tipa *backlink* (linkovi u okviru stranica koji upućuju na druge stranice) i na osnovu njih je određivao koliko su bitne pojedine stranice na Internetu. Projekat je dobio naziv *Backrub*, upravo zbog načina rada ovog mehanizma.

Želeli su da „organizuju sve informacije sveta i da ih učine univerzalno pristupačnim i korisnim“ i u tom duhu su *Backrub* preimenovali u *Google*, što predsatvlja matematički izraz za broj 1 koji je praćen sa 100 nula (1 googol).

Osnova *Google* mehanizma pretrage se zasniva na *PageRank* algortimu, koji koristi opisani koncpet *Backrub* mehanizma. Drugim rečima, ovaj algoritam istražuje koliko drugih sajtova referencira neku stranicu da bi odredio koliko je ta stranica zaista relevantna.

Naročitu pažnju velikog broja kompanija je privukao *Google* algoritam za reklame. Algoritam koristi reči koje korisnik unese tokom pretrage kako bi oforimo skup reklama koje mu treba prikazati [5][7][9].

Kada je potrebno nešto pretražiti, ljudi će za to obično reći da je to potrebno „izguglati“, što već dovoljno govori o tome gde se danas nalazi Google i njihov mehanizam pretrage. Pored ovog mehanizma, među popularnim mehanizmima su još i: *Microsoft Bing*, *Yahoo*, *Baidu* i drugi.

# Značaj i primena

Na fakultetima širom sveta prolazi i smenjuje se bezbroj studenata. Svaki od njih je tokom svojih studentskih dana uradio barem nekoliko projekata, što na zahtev predmetnog profesora, što iz radoznalosti i želje za unapređivanjem. Iako različiti, jedna stvar je zajednička svim studentima, a to je da uvek pitaju za savete i iskustva nekog od svojih starijih kolega. Iz tih razloga su tokom godina nastajali različiti *Github* repozitorijumi, *DropBox* linkovi, kompresovani folder i slično, puni materijala uz pomoć kojih je lakše položiti ispite.

Ideja podele materijala novim generacijama jeste vrlo dobra i odlično se pokazala kada je u pitanju teorijski deo ispita. Imajući to u vidu, nameće se jedno vrlo prosto pitanje, šta se dešava sa projektima koje su starije kolege radile? Projekti ne treba da se recikliraju godinama, zbog čega profesori i asistenti svake godine smišljaju nove za nove generacije studenata i time već urađeni projekti postaju u dovoljnoj meri neupotrebijivi.

Međutim, svaki od tih projekata je prilikom izrade imao po neki problem (*eng. bug*), a dosta njih je imalo slične ili iste probleme. Studenti koji su radili na njima su možda potrošili malo više vremena dok su pronašli adekvatno rešenje u moru informacija koje se danas nalaze na Internetu. Iz generacije u generaciju, studenti imaju slične probleme prilikom izrade svojih projekata bez uvida u to da li je neko od starijih kolega imao sličan problem, ali i bez načina kako ga je rešio.

Jedan od potencijalnih rešenja koje se koristi trenutno jeste da se uvede po 1 dokument sa pitanjima i odgovorima (*eng. QA sheet*) za svaku generaciju i svaki predmet. U okviru ovog dokumenta, studenti bi opisivali problem na koji su naišli i onda bi čekali odgovor nekog od asistenata. Ovo rešenje je dobro na generacijskom nivou, jer su onda svi studenti iste generacije u mogućnosti da vide odgovor na problem sa kojim se susreo neki od njihovih kolega. Glavni nedostatak ovakvog pristupa jeste što se taj dokument briše i pravi novi, za svaki predmet i za svaku generaciju ponovo.

Umesto dokumenata sa pitanjima i odgovorima, mnogo bolje rešenje je da postoji sajt koji bi obuhvatio sva implementaciona i diskusiona pitanja studenata, nezavisno od predmeta i tehnologija. Na ovakav način se postiže:

* Svaki student može da pogleda bilo koje pitanje i njegove odgovore, bez obzira na to kada je postavljeno.
* Studenti, asistenti i profesori mogu da odgovore na postavljena pitanja. Na takav način student koji je postavio pitanje brže dolazi do odgovora koji mu je potreban, a asistenti i profesori postaju rasterećeniji od gomile pitanja koje bi inače dobili.
* Najčešći problemi koji se javljaju tokom izrade projekata će vrlo brzo imati odgovore i neće biti potrebe za dupliranjem istih pitanja.
* Studenti se navikavaju da rešenje koje su pronašlo prilagode svom kodu.

## Postojeća rešenja

### *Stack Exchange* i *Stack Overflow*

*Stack Overflow* je sajt, zamišljen da predstavlja biblioteku koja bi umesto knjiga na svojim policama imala sva moguća pitanja vezana za programiranje. Detaljni i najkvalitetniji odgovori bi se nalazili na njihovim stranicama. Programeri širom sveta su upravo oni koji omogućuju da se ovaj cilj postigne, pomažući potpunim strancima na Internetu, nudeći im odgovore za opisane probleme [10].

*Stack Overflow* je samo jedan od 173 sajta koja se nalaze u okviru *Stack Exchange* mreže, koju na mesečnom nivou poseti preko 100 miliona ljudi. Trenutno predstavlja najmoćniju mrežu za pronalaženje odgovora, sa *Stack Overflow* kao vodećim sajtom. [11].

### *Discuss The Elastic Stack*

*Discuss The Elastic Stack* predstavlja skup „diskusionih foruma za *Elasticsearch*, *Beats*, *Logstash*, *Kibana*, *Elastic Cloud* i ostale proizvode u okviru *Elastic* ekosistema“ [12].

Slično kao i *Stack Overflow*, ovaj sajt je namenjen za rešavanje problema nastalih prilikom implementacije rešenja, koja su usko vezana za *Elastic* domen.

### *GitHub community*

*GitHub community* je sajt namenjen *GitHub* korisnima, kao mesto na kome mogu da dobiju odgovor na postavljeno pitanje i da pritom nauče nešto novo ili čak dobiju inspiraciju za neki novi projekat [13].

### Ostalo

Osim navedenih sajtova, koriste se i Dev.to, Experts Exchange, Code Project i slični.

# *Elasticsearch* mehanizam

## Šta je *Elasticsearch* mehanizam i kako je nastao?

*Elasticsearch* je distribuirani, *open source* mehanizam za pretragu i analitiku, čiji temelj predstavlja *Apache Lucene* bibilioteke. Omogućava skladištenje, pretragu i analiziranje velike količine podataka u realnom vremenu.

*Elasticsearch*, danas jedan od najpopularnijih mehanizama pretrage, nastao je kada je Shay Banon, suosnivač *Elastic* kompanije, želeo da kreira mehanizam pretrage koji bi njegova žena mogla da koristi za svoje recepte za kuvanje [16].

## Struktura i način rada *Elasticsearch* mehanizma

### Indeksi

*Elasticsearch* koristi svoju ugrađenu NoSQL bazu koja podatke čuva kao JSON dokumente. Dokumenti predstavljaju osnovnu jedinicu podatak koju je moguće grupisati u indeks, koncept sličan tabeli kod relacionih baza podataka. Da bi se dokumenti grupisali u indeks, moraju imati slične karakteristike [14].

Postoje 2 tipa strukture podataka koje se koriste kod mehanizama pretrage u cilju čuvanja i organizovanja podataka:

* *forward index* - dokumenti se mapiraju na termine koje sadrže. To znači da se čuva lista svih reči koje se nalaze u svakom dokumentu. Ova struktura je dobra kada je u pitanju indeksiranje, pošto se svaka reč nadovezuje na prethodnu, ali s druge strane, nije previše efikasna kada je reč o pretrazi na osnovu termina. Na slici 5 se nalazi primer koji prikazuje kako se korišćenjem *forward index* strukture podataka smešta 5 dokumenata i njihov sadržaj. Ukoliko korisnik želi da pronađe sve dokumente u kojima se pominje reč „*Elasticsearch*“, mehanizam bi morao da prođe kroz sve dokumente kako bi korisniku vratio rezultat „dokument1, dokument4“.



Slika 5. Forward index

* *inverted index* - termini se mapiraju na dokumente u kojima se nalaze. To znači da za svaku reč postoji lista dokumenata u kojima se nalazi. Kod ovakve strukture je indeksiranje sporije, pošto je za svaku reč potrebno proveriti da li već postoji u indeksu, ali je pretraga veoma brza, iz razloga što se podaci već unapred čuvaju kao upiti (*eng. queries*). To omogućava da pretrage budu jako brze, čak i kada postoji veoma velika količina podataka. *Elasticsearch* mehanizam pretrage koristi upravo ovu strukturu za smeštanje svojih podataka. Na slici 6 se nalazi primer koji prikazuje kako se korišćenjem *inverted index* strukture podataka smešta 5 dokumenata i njihov sadržaj. Ukoliko korisnik želi da pronađe sve dokumente u kojima se pominje reč „Elasticsearch“, mehanizam odmah može da vrati rezultat „dokument1, dokument4“, upravo iz razloga što se podaci čuvaju kao upiti [15].



Slika 6. Inverted index

### Cluster i node

# Zaključak

# Literatura

1. Capitol Technology University, „Alan Emtage Creator of Archie, the World's First Search Engine“*, https://www.captechu.edu/blog*, Artikal o kreatoru prvog mehnaizma pretrage.
2. LinkedIn, Archie interfejs, *https://media.licdn.com/dms/image*, Archie interfejs.
3. Stackscale Group Aire, „Archie, the first Internet search engine“*, https://www.stackscale.com/blog*, Artikal o prvom mehanizamu pretrage.
4. DBPedia, „History of Yahoo!“, *https://dbpedia.org*, Kratka istorija o nastanku Yahoo! mehanizma pretrage.
5. BusinessHistory, „Yahoo!: The first king of the internet“, *https://businesshistory.domain-b.com/focus*, Nastanak i razvijanje Yahoo! mehanizma pretrage
6. Chatri Sityodtong, „Lessons From Yahoo's Fall: $125b to $5b“, *https://chatrisityodtong.com/blog/entrepreneurship*, Artikal o tome kako je Google zamenio Yahoo!
7. Google, „From the garage to the Googleplex“, *https://about.google*, Nastanak Google kompanije
8. Vecteezy, *https://static.vecteezy.com*, Google interfejs
9. Search Engine Land, „A guide to Google: Origins, history and key moments in search“, *https://searchengineland.com/guide*, Istorija Google kompanije i njihovog mehanizma pretrage.
10. Stack Overflow, „Welcome to Stack Overflow“, https://stackoverflow.com, Upoznavanje sa Stack Overflow sajtom
11. Stack Exchange, „The world's largest programming community is growing“, https://stackexchange.com, O Stack Exchange mreži
12. Elastic, „About Discuss the Elastic Stack“, https://discuss.elastic.co, O Discuss the Elastic Stack sajtu
13. GitHub Communitz, „Overview“, https://github.com/community, Opis čemu je namenjen GitHub Community
14. Jay Gopalakrishnan, „Elasticsearch: What It Is, How It Works, And What It's Used For“, *https://www.knowi.com/blog*, Informacije o tome šta je i kako radi Elasticsearch
15. GeeksForGeeks, „Difference between Inverted Index and Forward Index“, https://www.geeksforgeeks.org, Opis razlike između inverted i forward index strukture podataka sa primerima
16. Elastic, „Recipes, Elasticsearch. More recipes, morphood + Elasticsearch“, *https://www.elastic.co/elasticon*, Kratak pregled za Webinar

# Dodatak A

Ovaj deo imaju samo neki radovi, većina radova ih nema.

To su obično elementi koji su preveliki da se nađu u samom sadržaju rada. Na primer: predugačke tabele, prevelike slike, predugačak kod.

# Dodatak B

# Podaci o kandidatu

Kandidat Ime Prezime je rođen/a xxxx. godine u Gradu. Završio/la je srednju školu u Gradu, 20xx. godine. Fakultet Tehničkih Nauka u Novom Sadu je upisao/la 20xx. godine. Ispunio/ispunila je sve obaveze i položio/la je sve ispite predviđene studijskim programom sa prosečnom ocenom od xx.xx.

Ovo je generički tekst, može se menjati i može se postaviti fotografija.