

Iskanje in ekstrakcija podatkov s spleta - 3. naloga

Anei Makovec, 63150186

Maj 2020

1 Uvod

Za 3. nalogo sem implementiral iskanje poizvedb iz besedil s spletnih stranih z uporabo inverznega indeksa in sekvenčnega branja datotek. Za implementacijo inverznega indeksa sem uporabil podatkovno bazo SQLite, zato sem izvajanje razdelil na dva dela: predprocesiranje in gradnjo podatkovne baze ter samo iskanje poizvedbe. Programe sem napisal v programskem jeziku Python. Najprej bom predstavil podrobnosti obeh implementacij, nato bom analiziral zgrajeno podatkovno bazo, na koncu pa še pokomentiral in primerjal rezultate posamezne implementacije.

2 Implementacija

Za inverzni indeks sem napisal dva programa, in sicer *run-indexing.py*, ki zgradi podatkovno bazo, ter *run-sqlite-search.py*, ki izvede iskanje. Za sekvenčno iskanje pa sem napisal le en program (*run-basic-search.py*), ki izvede predprocesiranje in iskanje hkrati. Obe implementaciji uporabljata funkcije iz datoteke *preproc.py*, ki sem jo sam napisal, da bi vsaka uporabljala enako predprocesiranje. Ta datoteka vključuje štiri funkcije:

- *preproces(text)* - vrne seznam terk v formatu (*indeks, beseda*), kjer *indeks* predstavlja mesto besede v podanem besedilu *text*, *beseda* pa predstavlja besedo iz besedila, ki ni stop-beseda in ki je bila normalizirana v majhne črke.
- *tokenize_text(text)* - vrne seznam, ki predstavlja tokenizirano celotno besedilo.
- *index_words(words, text)* - sprejme seznam, ki ga vrne funkcija *preprocess* kot *words*, ter seznam, ki ga vrne funkcija *tokenize_text* kot *text*, in vrne dva slovarja: slovar, ki vsebuje število (frekvenco) in vse indekse pojavitev določene besede v besedilu, ter slovar, ki vsebuje izsek besedila na določenemu indeksu.
- *get_snippet(text, index)* - sprejme seznam, ki ga vrne funkcija *tokenize_text*, in vrne izsek besedila na indeksu *index*.

Za tokenizacijo besedila se uporabi Pythonova knjižnica *nlTK*.

2.1 Inverzni indeks

2.1.1 Predprocesiranje in izgradnja podatkovne baze

Kot sem že omenil, se predprocesiranje pri implementaciji inverznega indeksa izvede s programom *run-indexing.py*. Ta zgradi podatkovno bazo (*inverted-index.db*), ki jo sestavljajo tri tabele. Prvi dve sta enaki, kot sta definirani v navodilih, tretjo pa sem dodal za hrambo izsekov besedil. Podatkovna baza torej vsebuje sledeče tabele:

- **IndexWord** - vsebuje seznam vseh besed, ki jih program sprocesira.
- **Posting** - vsebuje zapis o frekvenci in indeksu pojavitve določene besede v določeni datoteki.
- **Snippets** - vsebuje izseke besedil iz določene datoteke na določenem indeksu.

Program se torej, s pomočjo funkcij iz *preproc.py*, sprehodi po vseh datotekah, iz njihovega telesa (HTML značke *body*) prebere besedilo, ga spredprocesira in pridobi seznam s frekvenkami in indeksi besed ter seznam z izseki besedila, na koncu pa vse te podatke zapiše v podatkovno bazo.

2.1.2 Iskanje z inverznim indeksom

Program za iskanje poizvedb z inverznim indeksom je napisan v datoteki *run-sqlite-search.py*. Ta prebere vhodno poizvedbo, jo spredprocesira z uporabo funkcije *preprocess* in nato v podatkovni bazi poišče vse zapise iz tabele *Posting*, ki ustrezajo besedam iz poizvedbe. Zapisi vseh besed iz poizvedbe so združeni in urejeni že v poizvedbi, ki se opravi nad podatkovno bazo. Vrnjen seznam zapisov torej že vsebuje seštevek frekvenc in združene indekse pojavitev za vse besede iz poizvedbe v enaki datoteki, hkrati pa so ti tudi urejeni padajoče po številu frekvenc. Sedaj program samo še za vsako datoteko, v kateri so se pojavile besede iz poizvedbe, iz podatkovne baze pridobi izseke besedila na pridobljenih indeksih in jih združi. Celoten rezultat nato izpiše na standarden izhod v formatu, ki je specificiran v navodilih. Spodaj je podanih nekaj primerov izhodov opravljenih poizvedb, pri čemer je prikazanih le prvih nekaj rezultatov.

Results for a query: "predelovalne dejavnosti"

Results found in 0.1087s.

Frequencies	Document	Snippet
1291	evem.gov.si/evem.gov.si.371.html	... iskanje ustrezne šifre dejavnosti/storitve in informacij ... pogojih za opravljanje dejavnosti. V iskalnik špiite ... 645 od 645 dejavnosti Izpisanih je od ... Izpisanih je od dejavnosti A KMETIJSTVO IN ... pogojih za opravljanje dejavnosti: Pridelava semenskega materiala ...
75	evem.gov.si/evem.gov.si.377.html	... Defektolog v zdravstveni dejavnosti Dekan oziroma direktor ... Dietetik v zdravstveni dejavnosti Dimnikar Diplomirana medicinska ... I v zdravstveni dejavnosti Laboratorijski sodelavec II ... II v zdravstveni dejavnosti Laboratorijski tehnik Ladijski ...
40	podatki.gov.si/podatki.gov.si.340.html	... KALAN-NOSILEC DOPOLNILNE DEJAVNOSTI NA KMETIJI BREGAR ... šport CENTER INTERESNIH DEJAVNOSTI PTUJ CENTER JUDOVSKO ... ŠOLSKIH IN SOBOLSKIH DEJAVNOSTI Center urbane kulture ... in druge zdravstvene dejavnosti, d.o.o. DENTIM zobozdravstvo ... DERMA SAN, zdravstvene dejavnosti, prodaja, storitve, d.o.o ...
38	evem.gov.si/evem.gov.si.452.html	... Moj e-DEM >eDEM >Dejavnosti Druge storitvene dejavnosti ... >Dejavnosti Druge storitvene dejavnosti, druge ščnerazvrene (96.090 ... 96.090) Druge storitvene dejavnosti, druge ščnerazvrene (96.090 ... SKD šifra zajema dejavnosti in storitve, za ...
31	evem.gov.si/evem.gov.si.653.html	... Dovoljenje za opravljanje dejavnosti specializirane prodajalne z ... radijske ali televizijske dejavnosti Dovoljenje za izvajanje ... za izvajanje sevalne dejavnosti Dovoljenje za izvajanje ... za izvajanje sevalne dejavnosti Dovoljenje za izvajanje ...

Results for a query: "trgovina"

Results found in 0.0574s.

Frequencies	Document	Snippet
364	evem.gov.si/evem.gov.si.371.html	... organizacij, gl. 46.110 trgovina na debelo s ... juh, gl. 10.890 trgovina na debelo z ... ipd., gl. 10.890 trgovina na debelo s ... jedmi, gl. 46.380 trgovina na drobno s ... ščskladjenje nevarnih kemikalij Trgovina na debelo z ... z nevarnimi kemikalijami Trgovina na drobno z ...
96	evem.gov.si/evem.gov.si.651.html	... Druga govedoreja Druga trgovina na drobno v ... specializiranih prodajalnah Druga trgovina na drobno v ... nespecializiranih prodajalnah Druga trgovina na drobno v ... z živili Druga trgovina na drobno zunaj ... čNepremijsko posredovanje Nespecializirana trgovina na debelo Nespecializirana ...
92	evem.gov.si/evem.gov.si.21.html	... e-DEM >eDEM >Podroja Trgovina Tu boste šnali ... Seznam dejavnosti Druga trgovina na drobno v ... nespecializiranih prodajalnah Druga trgovina na drobno zunaj ... žtrnic(47.990) Nespecializirana trgovina na debelo Trgovina ... trgovina na debelo Trgovina na drobno na ...
82	podatki.gov.si/podatki.gov.si.340.html	... d.o.o. A DENT, trgovina in storitve, d.o.o ... d.o.o. ADRIA INVESTICIJE trgovina, šposrednitvo, storitve in ... storitve d.o.o. AHATSERVIS trgovina in storitve, d.o.o ... žvzdrevanje d.o.o. ALBA trgovina in proizvodnja, d.o.o ... proizvodnja, storitve in trgovina d.o.o. ALMA MATER ...
13	evem.gov.si/evem.gov.si.623.html	... e-DEM >eDEM >Dejavnosti Trgovina na debelo z ... izdelki široke porabe Trgovina na debelo z ... porabe Sem spada: trgovina na debelo z ... plutovinastimi izdelki ipd. trgovina na debelo s ... in deli zanja trgovina na debelo s ... ščpotrebinami , knjigami, časopisi trgovina na debelo z ...

Results for a query: "social services"

Results found in 0.0011s.

Frequencies	Document	Snippet
5	e-uprava.gov.si/e-uprava.gov.si.9.html	... culture Labour, retirement Social services, health, death ... Labour, retirement Social services, health, death Taxes ... employment relationship etc.? Social services, health, death ... relationship etc.? Social services, health, death How ... I obtain financial social assistance? How do ...
5	e-uprava.gov.si/e-uprava.gov.si.45.html	... culture Labour, retirement Social services, health, death ... Labour, retirement Social services, health, death Taxes ... employment relationship etc.? Social services, health, death ... relationship etc.? Social services, health, death How ... I obtain financial social assistance? How do ...
1	podatki.gov.si/podatki.gov.si.340.html	... recreation and spa services ltd. TERME MARIBOR ...
1	evem.gov.si/evem.gov.si.661.html	... Records and Related Services (AJPES) and the ...
...		

Results for a query: "Slovenija"

Results found in 0.0975s.

Frequencies	Document	Snippet
65	podatki.gov.si/podatki.gov.si.340.html	... NACIONALNI KOMITE PIARC SLOVENIJA, giz; v šanglekem ... PARK ŠKOCJANSKE JAME, Slovenija Partim, čfinanna in ... informiranje, d.o.o. RADIOTELEVIZIJA SLOVENIJA javni zavod, Ljubljana ... v likvidaciji REPUBLIKA SLOVENIJA REPUBLIKA SLOVENIJA, MINISTRSTVO ...
14	podatki.gov.si/podatki.gov.si.424.html	... in čstatistinih regijah, Slovenija, letno 55 ogledov ... špokodbe in spolu, Slovenija, letno 40 ogledov ... šsplonih zobnih ambulantah, Slovenija, letno 32 ogledov ... šsplonih zobnih ambulantah, Slovenija,... Nadaljujte z branjem ... spolu in starosti, Slovenija, leto 2014 26 ...
14	podatki.gov.si/podatki.gov.si.414.html	... Podrobnosti Organizacija: REPUBLIKA SLOVENIJA, MINISTRSTVO ZA DELO ... ENAKE ŽMONOSTI REPUBLIKA SLOVENIJA, MINISTRSTVO ZA DELO ... 149 ogledov REPUBLIKA SLOVENIJA, MINISTRSTVO ZA DELO ... 123 ogledov REPUBLIKA SLOVENIJA, MINISTRSTVO ZA DELO ...
14	e-prostor.gov.si/e-prostor.gov.si.166.html	... 132332,08 1012,4 VZHODNA SLOVENIJA MM 10000 528342,785 ... 120209,87 370,67 VZHODNA SLOVENIJA MM 10000 540391,312 ... 81709,09 374,87 VZHODNA SLOVENIJA MM 10000 521281,03 ... 33984,42 491,7 VZHODNA SLOVENIJA MM 10000 408572,817 ... 55648,42 741,66 VZHODNA SLOVENIJA MM 10000 416475,961 ...
12	podatki.gov.si/podatki.gov.si.350.html	... izdana gradbena dovoljenja, Slovenija, čmeseno 828 ogledov ... osebnih avtomobilov, gospodinjstva, Slovenija, čveletno 222 ogledov ... SKP-08) in spolu, Slovenija, letno 311 ogledov ... šNajpogostejša imena prebivalcev, Slovenija, čveletno 109 ogledov ...
...		

Results for a query: "ministrstvo za finance"

Results found in 0.1327s.

Frequencies	Document	Snippet
613	evem.gov.si/evem.gov.si.371.html	... kmetijskih rastlin Vir: Ministrstvo za kmetijstvo, gozdarstvo ... oziroma Živil Vir: Ministrstvo za kmetijstvo, gozdarstvo ... oziroma Živil Vir: Ministrstvo za kmetijstvo, gozdarstvo ... kmetijskih rastlin Vir: Ministrstvo za kmetijstvo , gozdarstvo ...
58	podatki.gov.si/podatki.gov.si.340.html	... gozdarstvo in prehrana Finance in davki Gospodarstvo ... gozdarstvo in prehrana Finance in davki Gospodarstvo ... CMILETI BOJAN-SIZVRITELJ MINISTRSTVO ZA DELO, ŽDRUINO ... SLOVENIJE ZA DELO MINISTRSTVO ZA FINANCE MINISTRSTVO ... DELO MINISTRSTVO ZA FINANCE MINISTRSTVO ZA FINANCE ...
54	podatki.gov.si/podatki.gov.si.348.html	... gozdarstvo in prehrana Finance in davki Gospodarstvo ... gozdarstvo in prehrana Finance in davki Gospodarstvo ... Javni sektor(236) Finance in davki (106 ... 227) REPUBLIKA SLOVENIJA, MINISTRSTVO ZA DELO, ŽDRUINO ... ENAKE ŽMONOSTI(61) MINISTRSTVO ZA OKOLJE IN ...
51	podatki.gov.si/podatki.gov.si.311.html	... gozdarstvo in prehrana Finance in davki Gospodarstvo ... gozdarstvo in prehrana Finance in davki Gospodarstvo ... in prostor(418) Finance in davki (403 ... 88) REPUBLIKA SLOVENIJA, MINISTRSTVO ZA DELO, ŽDRUINO ... ENAKE ŽMONOSTI(61) MINISTRSTVO ZA OKOLJE IN ...
43	podatki.gov.si/podatki.gov.si.214.html	... gozdarstvo in prehrana Finance in davki Gospodarstvo ... gozdarstvo in prehrana Finance in davki Gospodarstvo ... Organizacije Podrobnosti Organizacija: MINISTRSTVO ZA FINANCE MINISTRSTVO ... Organizacija: MINISTRSTVO ZA FINANCE MINISTRSTVO ZA FINANCE ...
...		

Results for a query: "predsednik zdrave"

Results found in 0.0767s.

Frequencies	Document	Snippet
40	evem.gov.si/evem.gov.si.371.html	... jo v imenu ždrave podeli vlada, posebej ... jo v imenu ždrave podeli vlada, posebej ... varnosti in obrambe ždrave. 5. člen ZEKom-1 ... varnosti in obrambe ždrave. 5. člen ZEKom-1 ... varnosti in obrambe ždrave. 5. člen ZEKom-1 ...
6	e-prostor.gov.si/e-prostor.gov.si.57.html	... Kontakt Ždravne ustanove----- Predsednik vlade Vlada RS ... Vlada RS Ministrstva Predsednik RS Ždravni zbor ... preko meja šnae ždrave. Razvoj novih satelitskih ... na celotnem ozemlju ždrave, in sicer v ... za žpreteni del ždrave zagotavlja čtonost transformacije ...
6	e-prostor.gov.si/e-prostor.gov.si.24.html	... Kontakt Ždravne ustanove----- Predsednik vlade Vlada RS ... Vlada RS Ministrstva Predsednik RS Ždravni zbor ... preko meja šnae ždrave. Razvoj novih satelitskih ... na celotnem ozemlju ždrave, in sicer v ... za žpreteni del ždrave zagotavlja čtonost transformacije ...
6	e-prostor.gov.si/e-prostor.gov.si.13.html	... Kontakt Ždravne ustanove----- Predsednik vlade Vlada RS ... Vlada RS Ministrstva Predsednik RS Ždravni zbor ... koordinatnem sistemu sosednje ždrave. Pri evidenci ždravne ... koordinatnem sistemu sosednje ždrave, lahko pa so ... v katerem te ždrave mejijo na Republiko ...
5	evem.gov.si/evem.gov.si.47.html	... Vodenje >podjetja Podpora ždrave pri razvoju Podpora ... pri razvoju Podpora ždrave pri razvoju Ste ... oziroma s čpomogo ždrave ustanovljene organizacije. Podporno ... organizacije. Podporno okolje ždrave (čnefinanna podpora): SPIRIT ... pri delu Podpora ždrave pri razvoju Odgovorno ...
...		

2.2 Iskanje s sekvenčnim branjem datotek

Pri implementaciji iskanja s sekvenčnim branjem datotek se predprocesiranje opravi v enakem programu, kot nato iskanje (*run-basic-search.py*), in sicer z uporabo enakih funkcij, kot pri implementaciji z inverznim indeksom. Razlika je v tem, da si program vse pridobljene podatke shrani v pomnilnik (in ne v podatkovno bazo) ter da si shrani le podatke o besedah, ki so v poizvedbi. Program se tako sprehodi po vseh datotekah, iz njih pridobi besedilo (iz HTML značke *body*), ga obdela s funkcijama *preprocess*, *tokenize_text* in *index_words* ter združi in shrani frekvenco in izseke besedila za besede iz poizvedbe. Na koncu pa še izpiše vse pridobljene rezultate na standardni izhod v formatu, ki je specficiran v navodilih. Spodaj je podanih še nekaj primerov izhodov opravljenih poizvedb, pri čemer je prikazanih le prvih nekaj rezultatov.

Results for a query: "predelovalne dejavnosti"

Results found in 60.2346s.

Frequencies	Document	Snippet
1291	evem.gov.si/evem.gov.si.371.html	... iskanje ustrezne šifre dejavnosti/storitve in informacij ... pogojih za opravljanje dejavnosti. V iskalnik švpiite ... 645 od 645 dejavnosti Izpisanih je od ... Izpisanih je od dejavnosti A KMETIJSTVO IN ... pogojih za opravljanje dejavnosti. Pridelava semenskega materiala ...
75	evem.gov.si/evem.gov.si.377.html	... Defektolog v zdravstveni dejavnosti Dekan oziroma direktor ... Dietetik v zdravstveni dejavnosti Dimnikar Diplomirana medicinska ... I v zdravstveni dejavnosti Laboratorijski sodelavec II ... II v zdravstveni dejavnosti Laboratorijski tehnik Ladijski ...
40	podatki.gov.si/podatki.gov.si.340.html	... KALAN-NOSILEC DOPOLNILNE DEJAVNOSTI NA KMETIJI BREGAR ... šport CENTER INTERESNIH DEJAVNOSTI PTUJ CENTER JUDOVŠKE ... ŠOLSKIH IN ŠOBOLSKIH DEJAVNOSTI Center urbane kulture ... in druge zdravstvene dejavnosti, d.o.o. DENTIM zobozdravstvo ...
38	evem.gov.si/evem.gov.si.452.html	... DERMA SAN, zdravstvene dejavnosti, prodaja, storitve, d.o.o. ...
31	evem.gov.si/evem.gov.si.653.html	... Moj e-DEM >DEM >Dejavnosti Druge storitvene dejavnosti ... >Dejavnosti Druge storitvene dejavnosti, druge ščnerazvrene (96.090 ... 96.090) Druge storitvene dejavnosti, druge ščnerazvrene (96.090 ... SKD šifra zajema dejavnosti in storitve, za ...
... Dovoljenje za izvajanje dejavnosti specializirane prodajalne z ... radijske ali televizijske dejavnosti Dovoljenje za izvajanje ... za izvajanje sevalne dejavnosti Dovoljenje za izvajanje ... za izvajanje sevalne dejavnosti Dovoljenje za izvajanje ...

Results for a query: "trgovina"

Results found in 61.5769s.

Frequencies	Document	Snippet
364	evem.gov.si/evem.gov.si.371.html	... organizacij, gl. 46.110 trgovina na debelo s ... juh, gl. 10.890 trgovina na debelo z ... ipd., gl. 10.890 trgovina na debelo s ... jedmi, gl. 46.380 trgovina na drobno s ... ščSkladienje nevarnih kemikalij Trgovina na debelo z ... z nevarnimi kemikalijami Trgovina na drobno z ...
96	evem.gov.si/evem.gov.si.651.html	... Druga govedoreja Druga trgovina na drobno v ... specializiranih prodajalnah Druga trgovina na drobno v ... nespecializiranih prodajalnah Druga trgovina na drobno v ... z živili Druga trgovina na drobno zunaj ... čNepremijsko posredovanje Nespecializirana trgovina na debelo Nespecializirana ...
92	evem.gov.si/evem.gov.si.21.html	... e-DEM >DEM >Podroja Trgovina Tu boste šnali ... Seznam dejavnosti Druga trgovina na drobno v ... nespecializiranih prodajalnah Druga trgovina na drobno zunaj ... žtrnic(47.990) Nespecializirana trgovina na debelo Trgovina ... trgovina na debelo Trgovina na drobno na ...
82	podatki.gov.si/podatki.gov.si.340.html	... d.o.o. A DENT, trgovina in storitve, d.o.o. ... d.o.o. ADRIA INVESTICIJE trgovina, šposrednitvo, storitve in ... storitve d.o.o. AHATSERVIS trgovina in storitve, d.o.o. ... žvzdrevanje d.o.o. ALBA trgovina in proizvodnja, d.o.o. ... proizvodnja, storitve in trgovina d.o.o. ALMA MATER ...
13	evem.gov.si/evem.gov.si.623.html	... e-DEM >DEM >Dejavnosti Trgovina na debelo z ... izdelki široke porabe Trgovina na debelo z ... porabe Sem spada: trgovina na debelo z ... plutovinastimi izdelki ipd. trgovina na debelo s ... in deli zanja trgovina na debelo s ... ščpotrebinami , knjigami, časopisi trgovina na debelo z ...
...

Results for a query: "social services"

Results found in 80.0552s.

Frequencies	Document	Snippet
5	e-uprava.gov.si/e-uprava.gov.si.9.html	... culture Labour, retirement Social services, health, death ... employment relationship etc.?
5	e-uprava.gov.si/e-uprava.gov.si.45.html	... culture Labour, retirement Social services, health, death ... employment relationship etc.?
1	evem.gov.si/evem.gov.si.661.html	... Records and Related Services (AJPEs) and the ...
1	podatki.gov.si/podatki.gov.si.340.html	... recreation and spa services ltd. TERME MARIBOR ...

Results for a query: "Slovenija"

Results found in 62.4983s.

Frequencies	Document	Snippet
-------------	----------	---------

```

-----
65      podatki.gov.si/podatki.gov.si.340.html ... NACIONALNI KOMITE PIARC SLOVENIJA, giz; v šanglekem ... PARK ŠKOCJANSKE JAME, Slovenija
Partim, čfinanna in ... informiranje, d.o.o. RADIOTELEVIZIJA SLOVENIJA javni zavod, Ljubljana ... v likvidaciji REPUBLIKA SLOVENIJA REPUBLIKA SLOVENIJA,
MINISTRSTVO ...
14      e-prostor.gov.si/e-prostor.gov.si.166.html ... 132332,08 1012,4 VZHODNA SLOVENIJA MM 10000 528342,785 ... 120209,87 370,67 VZHODNA
SLOVENIJA MM 10000 540391,312 ... 81709,09 374,87 VZHODNA SLOVENIJA MM 10000 521281,03 ... 33984,42 491,7 VZHODNA SLOVENIJA MM 10000 408572,817 ...
55648,42 741,66 VZHODNA SLOVENIJA MM 10000 416475,961 ...
14      podatki.gov.si/podatki.gov.si.414.html ... Podrobnosti Organizacija: REPUBLIKA SLOVENIJA, MINISTRSTVO ZA DELO ... ENAKE ŽMONOSTI
REPUBLIKA SLOVENIJA, MINISTRSTVO ZA DELO ... 149 ogledov REPUBLIKA SLOVENIJA, MINISTRSTVO ZA DELO ... 123 ogledov REPUBLIKA SLOVENIJA, MINISTRSTVO ZA
DELO ... 146 ogledov REPUBLIKA SLOVENIJA, MINISTRSTVO ZA DELO ...
14      podatki.gov.si/podatki.gov.si.424.html ... in čstatistinih regijah, Slovenija, letno 55 ogledov ... špokodbe in spolu, Slovenija, letno
40 ogledov ... šsplonih zobnih ambulantah, Slovenija, letno 32 ogledov ... šsplonih zobnih ambulantah, Slovenija,... Nadaljujte z branjem ... spolu in
starosti, Slovenija, leto 2014 26 ...
12      podatki.gov.si/podatki.gov.si.350.html ... izdana gradbena dovoljenja, Slovenija, čmeseno 828 ogledov ... osebnih avtomobilov,
gospodinjstva, Slovenija, čveletno 222 ogledov ... SKP-08) in spolu, Slovenija, letno 311 ogledov ... šNajpogostejša imena prebivalcev, Slovenija,
čveletno 109 ogledov ...
...

```

Results for a query: "ministrstvo za finance"

Results found in 63.6538s.

Frequencies	Document	Snippet
613	evem.gov.si/evem.gov.si.371.html	... kmetijskih rastlin Vir: Ministrstvo za kmetijstvo, gozdarstvo ... oziroma Živil Vir: Ministrstvo za kmetijstvo, gozdarstvo ... oziroma Živil Vir: Ministrstvo za kmetijstvo, gozdarstvo ...
58	podatki.gov.si/podatki.gov.si.340.html	... gozdarstvo in prehrana Finance in davki Gospodarstvo ... gozdarstvo in prehrana Finance in davki Gospodarstvo ... DELO MINISTRSTVO ZA FINANCE MINISTRSTVO ZA FINANCE ... FINANCE MINISTRSTVO ZA FINANCE ČFINANNA UPRAVA REPUBLIKE ... SLOVENIJE MINISTRSTVO ZA FINANCE UPRAVA REPUBLIKE SLOVENIJE ...
54	podatki.gov.si/podatki.gov.si.348.html	... gozdarstvo in prehrana Finance in davki Gospodarstvo ... gozdarstvo in prehrana Finance in davki Gospodarstvo ... Javni sektor(236) Finance in davki (106 ... 35) MINISTRSTVO ZA FINANCE (30) MINISTRSTVO ZA ... 20) MINISTRSTVO ZA FINANCE ČFINANNA UPRAVA REPUBLIKE ...
51	podatki.gov.si/podatki.gov.si.31.html	... gozdarstvo in prehrana Finance in davki Gospodarstvo ... gozdarstvo in prehrana Finance in davki Gospodarstvo ... in prostor(418) Finance in davki (403 ... 30) MINISTRSTVO ZA FINANCE (30) JAVNA AGENCIJA ... 20) MINISTRSTVO ZA FINANCE ČFINANNA UPRAVA REPUBLIKE ...
43	podatki.gov.si/podatki.gov.si.214.html	... gozdarstvo in prehrana Finance in davki Gospodarstvo ... gozdarstvo in prehrana Finance in davki Gospodarstvo ... Organizacija: MINISTRSTVO ZA FINANCE MINISTRSTVO ZA FINANCE ... FINANCE MINISTRSTVO ZA FINANCE čVkljui zbirke čherinskih ... zbirke ŽDravni organi Finance in davki čProraun ...
...		

Results for a query: "predsednik ždrave"

Results found in 62.6335s.

Frequencies	Document	Snippet
40	evem.gov.si/evem.gov.si.371.html	... jo v imenu ždrave podeli vlada, posebej ... jo v imenu ždrave podeli vlada, posebej ... varnosti in obrambe ždrave. 5. člen ZEKom-1 ... varnosti in obrambe ždrave. 5. člen ZEKom-1 ...
6	evem.gov.si/evem.gov.si.371.html	... jo v imenu ždrave podeli vlada, posebej ... jo v imenu ždrave podeli vlada, posebej ... varnosti in obrambe ždrave. 5. člen ZEKom-1 ... varnosti in obrambe ždrave. 5. člen ZEKom-1 ...
6	Ministrstva Predsednik RS ŽDravni zbor Kontakt ŽDravne ustanove----- Predsednik vlade Vlada RS ... Vlada RS
6	Ministrstva Predsednik RS ŽDravni zbor Kontakt ŽDravne ustanove----- Predsednik vlade Vlada RS ... Vlada RS
6	Ministrstva Predsednik RS ŽDravni zbor Kontakt ŽDravne ustanove----- Predsednik vlade Vlada RS ... Vlada RS
5	Ministrstva Predsednik RS ŽDravni zbor Kontakt ŽDravne ustanove----- Predsednik vlade Vlada RS ... Vlada RS
...		

3 Analiza podatkovne baze

V podatkovni bazi je indeksiranih 47254 besed. Podrobnosti podatkovne baze prikazujeta sliki 1 in 2.

Beseda	Frekvenca
podatkov	11048
slovenije	9928
republike	8572
dejavnosti	5564
podatki	4940

Slika 1: Besede z največjimi frekvencami.

Datoteka	Frekvenca
evem.gov.si/evem.gov.si.371.html	80959
podatki.gov.si/podatki.gov.si.340.html	27391
e-prostor.gov.si/e-prostor.gov.si.166.html	11163
e-prostor.gov.si/e-prostor.gov.si.218.html	4500
e-prostor.gov.si/e-prostor.gov.si.147.html	4191

Slika 2: Datoteke z največjimi frekvencami.

4 Primerjava rezultatov

Čas izvajanja posamezne implementacije glede na poizvedbo prikazuje slika 3.

Poizvedba	Čas[s] (inverzni indeks)	Čas[s] (sekvenčno branje datotek)
predelovalne dejavnosti	0.4708	60.2346
trgovina	0.0189	61.5769
social services	0.0011	63.7414
Slovenija	0.0975	62.4983
ministrstvo za finance	0.1327	63.6538
predsednik države	0.0767	62.6335

Slika 3: Čas izvajanja posamezne implementacije glede na poizvedbo.

Glede na pridobljene rezultate, lahko sklepam, da je implementacija z inverznim indeksom daleč hitrejša kot tista s sekvenčnim branjem datotek. Kar je tudi smiselno, saj je pri inverznemu indeksu potrebno le enkrat prebrati vse datoteke (ob gradnji podatkovne baze), medtem ko pa je potrebno pri sekvenčnem branju datotek le-te prebrati ob vsakem iskanju. Vendar pa se velikost podatkovne baze, zaradi razširitve, da lahko hrani tudi izseke besedil, znatno poveča. Razširjena podatkovna baza je velika 176,7 MB, osnovna podatkovna baza pa le 49,2 MB. Če bi torej implementacijo z inverznim indeksom implementiral tako, da se v podatkovni bazi ne bi hranili vsi izseki besedil in bi program za iskanje moral ponovno odpreti datoteke in iz njih prebral izseke, bi prihranil na prostoru, vendar bi se čas izvajanja programa povečal.