

UNIVERZA V LJUBLJANI  
EKONOMSKA FAKULTETA

**DIPLOMSKO DELO**

PRIHODNOST SPLETNIH ISKALNIKOV

Ljubljana, julij 2009

IVAN PRENDŽOV

## **IZJAVA**

Študent Ivan Prendžov izjavljam, da sem avtor tega diplomskega dela, ki sem ga napisal pod mentorstvom dr. Jake Lindiča, in dovolim objavo diplomskega dela na fakultetnih spletnih straneh.

V Ljubljani, julij 2009

Podpis: \_\_\_\_\_

# KAZALO

<b>UVOD .....</b>	<b>1</b>
<b>1 SVETOVNI SPLET IN SPLETNI ISKALNIKI .....</b>	<b>3</b>
<b>1.1 Splošno o internetu .....</b>	<b>3</b>
<b>1.2 Razvoj spletnih iskalnikov .....</b>	<b>4</b>
1.2.1 Seznami in direktoriji .....	5
1.2.2 Prvi iskalniki .....	6
<b>1.3 Delovanje iskalnika .....</b>	<b>7</b>
<b>2 KONCEPTUALNA DELITEV SPLETNIH ISKALNIKOV .....</b>	<b>8</b>
<b>2.1 Odprtokodni iskalniki .....</b>	<b>10</b>
<b>2.2 Semantični iskalniki .....</b>	<b>11</b>
<b>2.3 Socialni iskalniki .....</b>	<b>12</b>
<b>2.4 Vizualni iskalniki .....</b>	<b>15</b>
<b>2.5 Meta iskalniki .....</b>	<b>18</b>
<b>3 TRŽNI VODJE .....</b>	<b>18</b>
<b>3.1 Google .....</b>	<b>20</b>
<b>3.2 Najdi.si .....</b>	<b>25</b>
<b>3.3 Izzivalci .....</b>	<b>27</b>
<b>4 PRIHODNOST ISKALNIKOV .....</b>	<b>28</b>
<b>SKLEP .....</b>	<b>30</b>
<b>SEZNAM LITERATURE IN VIROV .....</b>	<b>32</b>



## UVOD

Svetovni splet ponuja veliko in vsak dan rastočo množico informacij in podatkov. Prav tako se dnevno veča število uporabnikov svetovnega spleta, ki na njem neprestano objavljajo nove informacije. Deskanje po internetu je v večini primerov namenjeno iskanju informacij, zabavi in v zadnjih letih vedno bolj povezovanju med ljudmi. Iskanje podatkov je vsak dan težje zaradi vedno novih virov informacij in vedno lažjega dostopa do informacij. V vsej tej množici informacij mora vsak posameznik vedeti, kje in kako iskati. Orodje, ki posamezniku pomaga najti določeno informacijo, je spletni iskalnik.

Spletni iskalniki želijo uporabniku ponuditi relevantne rezultate, vendar je pogosto ena informacija različno relevantna za različni spletni iskalnik in tudi različno relevantna za vsakega posameznika. Spletni iskalniki so zaradi izbire »prave« informacije postali zelo pomemben del svetovnega spleta in naših osebnih predstav o tem, katera informacija je prava. Bistvo vsakega iskalnika je tehnologija, ki poišče pravo informacijo, ki jo uporabnik išče.

Zadetki so bili pogosto zelo netočni, saj se je moral uporabnik med množico predlaganimi rezultati zelo potruditi, da je našel informacijo, ki jo je iskal. Revolucijo med iskalniki je naredil Google s svojo izjemno tehnologijo iskanja pravih podatkov. Uporabniki pa imamo še vedno mnogokrat težave pri iskanju pravih informacij in podatkov. Po drugi strani spletni iskalniki ponujajo veliko poslovnih možnosti tako za same spletne iskalnike kot uporabnike iskalnikov.

V diplomskem delu izhajam iz tega, da želi vsak spletni iskalnik biti čim boljši. Biti dober iskalnik pomeni, da so rezultati iskanja relevantni iskalni poizvedbi brez prikazovanja nebistvenih rezultatov. Če spletni iskalnik ne najde rezultatov za neko iskalno poizvedbo, bi moral ponuditi spremembe pri poizvedbi, da bo uporabnik našel, kar je iskal. Po drugi strani nam mora spletni iskalnik v primeru, če najde preveč informacij, znati pomagati izostriti našo poizvedbo tako, da bomo našli tisto, kar smo iskali. V diplomskem delu predpostavljam, da spletni iskalniki še vedno ne prikazujejo najboljših rezultatov na najboljši način (Baker, 2003).

Svetovni splet, osebni računalniki in vsebina na svetovnem spletu se vsakodnevno spreminjajo, zato se morajo dnevno spreminjati tudi spletni iskalniki (Bhatia, MPS & Khalid, Akshi Kumar, 2008).

Diplomskega dela sem se lotil z raziskovanjem različnih konceptov spletnih iskalnikov, ki so bili končnim uporabnikom v zadnjem času ponujeni v uporabo. V diplomskem delu se sprašujem, ali je možno, da postane nek novi spletni iskalnik z inovativnim konceptom bolj

uspešen od Googla, ki je danes daleč najbolj priljubljen spletni iskalnik, in mu odvzame znaten del obiskovalcev.

Namen diplomskega dela je predstaviti spletni iskalnik kot orodje za iskanje, opisati različne koncepte spletnih iskalnikov in prikazati rešitve za boljše iskanje.

Cilji diplomskega dela so:

- predstaviti bistvene lastnosti interneta, ki vplivajo na spletne iskalnike,
- predstaviti razvoj spletnih iskalnikov,
- med seboj primerjati različne vrste spletnih iskalnikov,
- predstaviti pomanjkljivosti trenutnih spletnih iskalnikov,
- raziskati, kateri koncepti skušajo reševati probleme trenutnih spletnih iskalnikov.

Pri pisanju diplomskega dela sem se opiral na strokovne vire in rezultate raziskav – sekundarne vire. Pri pisanju sem si pomagal tudi z različnimi članki, intervjuji in video konferencami oseb, ki imajo močan vpliv na razvoj svetovnega spleta. Za potrditev svojih trditev sem do neke mere preučil tudi delovanje nekaterih spletnih iskalnikov, predvsem Googla in Najdi.si, ki sta najbolj priljubljena iskalnika v Sloveniji. Uporabil sem tudi znanje, ki sem ga pridobil pri projektih, v katerih sem sodeloval v podjetju SRC.SI (eUprava, eVEM, ISPO), ter teoretično in praktično znanje, ki sem ga pridobil med študijem na poslovno-informacijski smeri Ekonomske fakultete.

V prvem poglavju diplomske naloge bom opisal pomen interneta in njegovo obliko, in sicer za lažje razumevanje spletnih iskalnikov, ki jih bom opisal v drugem delu, kjer bom predstavil razvoj in zgodovino spletnih iskalnikov do danes.

V tretjem delu bom konceptualno razdelil spletne iskalnike in pojasnil, katere probleme želi določen koncept reševati. Spletne iskalnike bom razdelil na smiselne skupine in jih analiziral.

Četrty del diplomskega dela bo namenjen tržnim vodjem in njihovim izzivalcem. Opredelil bom težave, ki pestijo obstoječe spletne iskalnike, in razložil, zakaj jih je treba izboljševati. Zadnji del diplomske naloge pa bom namenil prihodnosti spletnih iskalnikov. S pomočjo prednosti in slabosti različnih konceptov spletnih iskalnikov ter primerov tržnih vodij in njihovih izzivalcev bom skušal ugotoviti, kakšni bodo spletni iskalniki v prihodnosti.

# 1 SVETOVNI SPLET IN SPLETNI ISKALNIKI

Spletni iskalniki so del svetovnega spleta in brez njega ne morejo obstajati, zato bom najprej opredelil internet in svetovni splet.

## 1.1 Splošno o internetu

Internet imenujemo tudi informacijska avtocesta (angl. *Information Highway*), mreža (angl. *Net*), matrica (angl. *Matrix*) ali splet (angl. *Web*). Svetovni splet (angl. *World Wide Web* – WWW) je le uporabniku prijazen multimedijski del interneta (Burnett & Marshall, 2003, str. 1), ki povezuje dokumente, slike in druge vire, povezane s hipertekstovnimi povezavami (angl. *Hyperlink*) in URL-ji (angl. *Uniform Resource Locator*).

Tehnološko je internet javno dostopno računalniško omrežje, ki prenaša podatke v paketkih po standardnem internetnem protokolu (IP). Internet je mreža mrež, sestavljena iz množice lokalnih, akademskih, poslovnih in državnih mrež, ki skupaj nudijo informacije in servise.

Internet je pojem, ki označuje »komunikacijsko mrežo, ki povezuje mnogo različnih elektronskih tehnologij, ki so med seboj povezane ter komunicirajo in ne morejo več obstajati vsaka zase« (Burnett & Marshall, 2003, str. 2). Internet omogoča različne komunikacije uporabnikov, kot so na primer spletne klepetalnice (IRC, ICQ, MSN Messenger), novičarske skupine (Usenet), elektronska pošta (e-mail) in tako naprej (Burnett & Marshall, 2003, str 3).

Danes nekateri servisi in storitve, ki so bile še pred kratkim zelo priljubljene, niso več aktualni. Danes je aktualen svetovni splet druge generacije (angl. *Web 2.0*), ki pa je v svojem bistvu le moderen termin, saj je mnogo uveljavljenih in starih spletnih strani že vse od svojega nastanka »Web 2.0« (npr. amazon.com in craigslist.org). Druga generacija svetovnega spleta predstavlja drugačen način uporabe svetovne mreže in ne nekih tehničnih in drugih izboljšav. Druga generacija svetovnega spleta omogoča uporabnikom več kot le pridobivanje informacij po obstoječih dokumentih. Omogoča mrežno platformo (računalniško okolje), ki omogoča, da uporabnik uporablja aplikacije preko svojega brskalnika (O'Reilly, 2005).

Uporabniki sami dodajajo vrednost spletnim mestom s svojimi prispevki. Mnoge zelo popularne spletne strani omogočajo tudi socialno druženje (angl. *Social networking*). Zaradi vedno bolj enostavne uporabe želi vse več ljudi objavljati spletne vsebine, zato so danes zelo priljubljene blogi, Wikipedia ter spletna mesta, kot so YouTube, MySpace, FaceBook ... Druge

priljubljene storitve svetovnega spleta so še spletna telefonija (Skype) in tehnologija »vsak z vsakim« (angl. *Peer to Peer* ali P2P).

V nadaljevanju svoje diplomske naloge se bom dotaknil tudi termina Web 3.0, ki predstavlja hipotezo, kakšen bi lahko bil svetovni splet v prihodnosti.

Knjižnice imajo strukturirano bazo podatkov, v katero ročno vnašajo nove podatke (knjige in podatke o knjigah). Knjižničarji znajo zato hitro najti želeno knjigo ali knjige, ki jih iščejo na določeno temo. To lahko storijo zato, ker imajo o vsaki knjigi zapisan kratek povzetek, ključne besede, knjige so v katalogu po nekem določenem ključu. Vsaka knjiga v knjižnici je tudi ročno pregledana in vložena v katalog, kot je treba, z vsemi podatki o knjigi (Berry & Browne, 2005, str 1-2).

Strukturirani podatki so hranjeni elektronsko, kjer ima vsaka informacija pripet podatek o pomenu in obliki informacije. Podatki so hranjeni v podatkovnem modelu tako, da jih računalniški program lahko uporabi enostavno in učinkovito z matematičnimi in logičnimi koncepti (MG Rush, 2009).

Spletni iskalniki pa za razliko od knjižničarjev brskajo po internetu, ki je nestrukturirana baza podatkov, indeks spletnih iskalnikov (katalog) pa se osvežuje avtomatsko.

Svetovni splet je delno strukturirana baza podatkov, kjer so podatki med seboj semantično povezani, podobne vsebine so med seboj delno povezane, vendar pa vsebine niso enako označene (Wood, 2009).

Nestrukturirani podatki so podatki, ki so (ponavadi) v elektronski obliki in nimajo podatkovnega modela ali pa ga ni moč uporabiti z računalniškim programom (MG Rush, 2009).

## **1.2 Razvoj spletnih iskalnikov**

Spletni iskalnik je orodje, ki je namenjeno iskanju informacij po svetovnem spletu. Informacije pridobiva s pomočjo ključnih besed in jih prikaže v obliki povezav na spletne strani ali datoteke na svetovnem spletu. Informacije, ki jih spletni iskalnik prikaže kot rezultat, so spletne strani, dokumenti, slike in druge vsebine. Nekateri spletni iskalniki iščejo podatke tudi po »newsgroup« direktorijih, specifičnih bazah in drugih direktorijih, ki so prosto dostopni (Sullivan, 2002).



Podatki so shranjeni na vsem svetovnem spletu na različnih omrežjih in spletnih mestih. Iskalnik shrani le grobo ogrodje spletne strani (npr. do 100 kb HTML-kode). Iskalniki za pridobivanje informacij uporabljajo pajke ali črve (angl. *Crawlers*, *Spiders*). Pajek je spletni programček, ki se sprehaja po spletnih straneh. Na vsaki spletni strani obišče vse povezave, ki jih na njej najde. Spletne strani zapisuje v indeks (angl. *Index*). Na tak način lahko hitro objavi povezave v rezultatih do novih spletnih mest ali novih spletnih strani, rezultate iskanj pa hitro najde z iskanjem le po svojih indeks straneh.

Bistvo vsakega iskalnika je njegov algoritem. Algoritem določa, na kakšen način določi relevantnost rezultatov, in jih ponudi uporabniku.

Nagla rast interneta je ponudila ogromno bazo informacij in podatkov. Znanje je zato vsem lahko, vendar ne hitreje dostopno. Uporabnik mora sam znati uporabiti iskalna orodja ter priti do informacije, ki je resnično pravilna in relevantna. Za uporabnika je zato najpomembnejši izbor ključnih besed (Gerkeš, 2000).

### **1.2.1 Seznami in direktoriji**

Leta 1990 je bila ustvarjena prva spletna stran (<http://info.cern.ch/>). Do leta 1992 je bilo ustvarjenih 26 spletnih strani. Ob nastanku (1993) prvega množično uporabljenega spletnega brskalnika (NCSA Mosaic) je bila vsaka novonastala spletna stran objavljena na spletni strani »What's New« ([http://www.dejavu.org/prep\\_whatsnew.htm](http://www.dejavu.org/prep_whatsnew.htm)). V nadaljevanju je bila dnevno objavljena približno ena nova stran (Brin, 2009).

Začetek iskanja po svetovnem spletu so začeli sezname ali direktoriji. To so spletne strani, ki so jih običajno urejali in dopolnjevali entuziasti, ki so bili pogosto na svojem področju veliki strokovnjaki. Primer seznama je bil Yanoff List, ki je bil ustvarjen leta 1991. Seznam je bil urejen po abecedi, skupine pa ločene po kategorijah, kot so npr. rumene strani (Wired, 2009; Conntech Listserv Archives 2004, 2009).

Leta 1994 je na Inštitutu Jožefa Štefana nastal projekt Matkurja, ki je bil prvi spletni imenik v Sloveniji. Matkurja je uporabniku ponudila novo uporabniško izkušnjo, saj se je v tistem času internet v Sloveniji šele dodobra pojavil (Oblak, 2008, str. 9).

Ob koncu osemdesetih in v začetku devetdesetih let se je internet že zelo razširil. Količina podatkov na spletu je postajala prevelika, da bi se lahko zanesli le na sezname, ki jih ljudje ročno osvežujejo. Iskanje podatkov je postalo ozko grlo interneta.

## 1.2.2 Prvi iskalniki

Za prvi znani spletni iskalnik velja Archie. Na spletu je zaživel leta 1990, ustvaril pa ga je nek študent na Kanadski univerzi. Archie je shranjeval sezname vseh dostopnih datotek, ki so bile prosto dostopne preko različnih FTP-strežnikov. Iskalnik je za iskanje datotek uporabljal pajke. Uporabnik iskalnika Archie je poiskal mesto, kjer se je nahajala določena datoteka, ter si rezultate preko elektronske pošte poslal k sebi (Asadi, 2004).

V naslednjem letu je zaživel iskalnik z imenom Gopher, ki je že izvajal indeksiranja besedilnih datotek po strežnikih Gopher in uporabljal prve hiperpovezave, kot jih poznamo danes na svetovnem spletu. Uporabnik je v iskalnik vnesel iskano ključno besedo, Gopher pa je ponudil seznam rezultatov, ki so ustrezali. Uporabnik je nato ob želenem rezultatu s pritiskom na gumb enter prišel do Gopher naslova na strežniku, na katerem je bila iskana datoteka. Nato si je lahko uporabnik enako kot pri iskalniku Archie rezultate poslal na svoj elektronski naslov ali pa datoteko prebral na spletu. Tekstovne datoteke, po katerih je iskal Gopher, so kasneje postale spletne strani. Gopher so za iskanje uporabljali tudi iskalniki, kot so na primer Veronica in Jughead. Iskanja so se lotili podobno kot Archie, vendar so za iskanje uporabljali množico Gopherjevih strežnikov. Jughead se je za razliko od Veronice iskanja loti le po izbranih Gopherjevih strežnikih (Griffiths, 2006).

Spletni iskalnik, ki je že iskal informacije po svetovnem spletu, je Wandex (1993). Nato se je pojavili še Aliweb. Kasneje se je pojavil Jumpstation, ki je iskal le po naslovih spletnih dokumentov. Leta 1994 pa se je pojavil prvi spletni iskalnik, ki je uporabnikom že omogočal iskanje po ključnih besedah, ki se lahko pojavijo kjer koli v dokumentih. Imenoval se je WebCrawler. WebCrawler je bil tudi prvi spletni iskalnik, ki je postal širše znan v javnosti.

Kmalu potem se je pojavil Lycos, ki je bil tudi komercialno uspešen. V naslednjih letih se je pojavilo mnogo spletnih iskalnikov (npr. Excite, Inktomi, AltaVista). Pojavil se je tudi Yahoo!, ki ima še vedno svoj spletni imenik, v katerem so zapisane spletne strani, uporabnik pa jih lahko prebira. Leta 1995 se je pojavil Microsoftov iskalnik MSN (msn.com).

Leta 1996 pa se je pojavil tudi Google, ki je za razvrščanje spletnih strani po relevantnosti začel uporabljati tehnologijo, ki računa povezave med spletnimi stranmi ter daje večjo spletnim stranem, ki imajo več povezav z drugih spletnih mest. Google je še do danes ohranil minimalistično zasnovo, ki ni spletni portal kot večina drugih iskalnikov.

Po letu 2000, ko je večina uporabnikov že uporabljala enega od velikih spletnih iskalnikov, so novi spletni iskalniki želeli tržni delež hitro rastočega Googla. Primer iskalnika, ki je že znal personalizirati rezultate iskanj ter prikazati rezultate na bolj inovativen način, je a9 (a9.com).

V Sloveniji sta se pojavila spletna iskalnika Slowwwenia in Raziskovalec. V današnjem času pa je najbolj obiskan Najdi.si, ki je tudi sicer najbolj obiskana slovenska spletna stran. Najdi.si je iskalnik, ki išče izključno po slovenskih spletnih dokumentih (Najdi.si, 2008).

### 1.3 Delovanje iskalnika

Spletni iskalnik deluje po naslednjem vrstnem redu:

- spletno »plazenje« (angl. *web crawling*),
- indeksiranje (angl. *indexing*),
- iskanje (angl. *searching*).

Spletni iskalnik deluje na način, da shranjuje informacije o spletnih straneh, ki jih najde na svetovnem spletu. Spletne strani so poizvedene s pomočjo spletnih črvov ali pajkov. Vsebina spletnih strani je nato shranjena v indeks. Iskanje poteka tako, da s pomočjo ključnih besed, vpisanih v uporabniški vmesnik, naredimo poizvedbo, ki se izvede po indeksiranih dokumentih.

Spletni iskalnik kot rezultat iskanja ponudi najbolj relevantne spletne strani kot seznam s kratkim povzetkom spletne strani, ki navadno vsebuje naslov spletnih strani in nekaj besedila s spletne strani, ki se najbolj ujema s poizvedbo.

Seznam rezultatov je običajno urejen po relevantnosti, ki je določena z matematičnim algoritmom. Matematični algoritem pa se razlikuje od spletnega iskalnika do iskalnika. Algoritem se pri iskalnikih tudi pogosto spreminja, saj le tako lahko sledi trendom in razvoju.

Spletni iskalniki imajo lahko svoje lastne algoritme in svoj lasten indeks spletnih strani ali pa so odprtokodni in vanj lahko posega vsak (npr. [search.wikia.com](http://search.wikia.com)).

#### Iskalni črv

Iskalni črvi so aplikacije oziroma avtomatizirana skripta, ki obiskujejo spletne strani na svetovnem spletu in njihove povezave. Nove najdene spletne strani shranijo v indeks in osvežijo shranjene podatke o spletnih straneh, ki so v njem že zapisane.

Iskalni črvi spletne strani obiskujejo avtomatično in periodično. Iskalni črvi obišejo spletne strani, ki imajo večkrat novo vsebino (npr. novičarske portale), bolj pogosto kot spletne strani, ki imajo bolj statično vsebino.

Iskalni črvi omogočajo, da so spletni iskalniki osveženi in imajo v svojem indeksu najnovejše podatke o spletnih straneh.

## **Indeks**

Indeks je baza podatkov, v kateri iskalni črvi zbirajo in hranijo podatke z namenom, da lahko spletni iskalnik hitro najde pravo informacijo. Spletni iskalnik ob iskanju podatkov ne išče po vsem svetovnem spletu, ampak le po shranjenih podatkih o spletnih straneh v svojem indeksu. Spletni iskalniki v svoj spletni indeks zapisujejo celotno tekstovno datoteko kot tudi zvočne in slikovne datoteke.

Brez spletnega indeksa bi spletni iskalnik ob vsakem zahtevku po iskanju iskal po vsem spletu po vsebini vseh datotek. To bi zelo podaljšalo čas iskanja in obenem zahtevalo mnogo večje računalniške zmogljivosti.

## **Iskanje**

Z uporabniškega vidika se iskanje s pomočjo spletnega iskalnika izvede tako, da spletni iskalnik vpiše ključno besedo v vnosno polje ter pritisne na gumb »išči«.

Iskalnik pa v fazi iskanja primerja iskalno poizvedbo z indeksom in prikaže rezultate, ki se mu zdijo najbolj relevantni.

## **2 KONCEPTUALNA DELITEV SPLETNIH ISKALNIKOV**

Poznamo različne kategorije spletnih iskalnikov. Iskalnike lahko delimo glede na:

- spletno vsebino (horizontalni iskalniki, vertikalni iskalniki);
- tip informacij (forumi, blogi, multimedija, torrenti, elektronski naslovi, zemljevidi, cene, vprašanja in odgovori ...);
- model (odprtokodni iskalniki, semantični iskalniki, socialni iskalniki, meta iskalniki, vizualni iskalniki, iskalne naprave, namizni iskalniki, usenet);
- to, da temeljijo na nekem drugem iskalniku (Google, Yahoo!, Live Search, Ask ...).

Spletne iskalnike lahko v splošnem delimo glede na vsebino, ki jo spletni iskalnik išče. Iskalnike zato delimo na horizontalne in vertikalne iskalnike. Horizontalni spletni iskalniki so spletni iskalniki, ki informacije iščejo po najširšem svetovnem spletu. Vertikalni ali specialni iskalniki pa so iskalniki, ki iščejo znotraj določene niše, področja ali določenega spletnega mesta oziroma domene. Vertikalni iskalniki so namenjeni iskanju samo določenih vsebin ali pa iskanju določenega tipa informacij. Navpično ali vertikalno iskanje je torej iskanje po neki vsebini, ki jo takšen iskalnik pokriva (ki je omejen le na temo te vsebine) (Enge, 2007).

Primeri iskalnikov, ki iščejo le po določenih vsebinah, so spletni iskalniki za iskanje služb, zdravstvenih napotkov, nepremičnin, nakupovanja, ekonomije, iskalniki po znanstvenih bazah, zaprtih bazah, pravu, medicini, novicah, intranetu ...

Vertikalni iskalniki se od horizontalnih iskalnikov (Google, Yahoo! ...) razlikujejo tudi po tem, da njihovi iskalni črvi običajno iščejo po strukturirani bazi podatkov ali indeksu, ki ima boljše meta podatke o dokumentih.

Glavna prednost vertikalnih iskalnikov je, da za specifična iskanja ponudijo bolj relevantne rezultate. Primer, kjer običajni horizontalni iskalniki odpovedo, je iskanje informacij znotraj specifične dejavnosti ali branže, npr. informacijske tehnologije, kirurgije, računovodstva (Sullivan, 2002).

Če zobotehnik v horizontalni iskalnik vpiše ključno besedo za poizvedbo »keramika«, bo iskalnik ponudil mnogo rezultatov, ki niso niti približno takšni, kot bi si jih želel zobotehnik. Če pa bi zobotehnik vpisal enako ključno besedo v iskalnik, ki je specializiran za zobotehniko, bi dobil mnogo bolj relevantne rezultate.

Slabost vertikalnih iskalnikov pa je, da so nove spletne strani, ki so vključene v iskanje, pogosto dodane ročno, kar pomeni, da lahko preteče nekaj časa od nastanka spletne strani do vključitve strani v indeks spletnega iskalnika.

Primer iskalnikov, ki iščejo le po določenem tipu informacij, so npr. iskalniki slik, ki iščejo le po slikovni vsebini, ali pa video iskalniki, ki iščejo le po video vsebinah.

V svojem diplomskem delu se osredotočam na delitev iskalnikov glede na model oziroma glede na način delovanja v konceptualnem smislu. Ta delitev omogoča prikaz prednosti in pomanjkljivosti določenih iskalnikov ter možnega razvoja v prihodnosti.

## 2.1 Odprtokodni iskalniki

Odprtokodni iskalniki omogočajo, da uporabnik objavi spremembe in vnose (prispevke) za izboljšavo spletnega iskalnika. Odprtokodni iskalniki so običajno zastojni, zahtevajo le uporabo, ki je pod pogoji javne licence (angl. *GPL*) ali neke druge odprtokodne licenčne sheme. Smisel odprte kode je v tem, da lahko vsak uporabi programsko kodo oziroma spletni iskalnik, vendar mora z drugimi uporabniki deliti izboljšave in dodatne funkcionalnosti, ki jih naredi. Tehnična podpora pri odprtokodnih programih je navadno mogoča le preko drugih uporabnikov ali razvijalcev preko elektronske pošte, elektronskih konferenc ali preko plačljivih svetovalcev.

Primeri odprtokodnih iskalnikov so: Egothor ([egothor.org](http://egothor.org)), DataparkSearch Engine ([dataparksearch.org](http://dataparksearch.org)), ASPseek ([aspseek.org](http://aspseek.org)) in Nutch ([lucene.apache.org/nutch/](http://lucene.apache.org/nutch/)).

Prednosti odprtokodnih iskalnikov so enake kot prednosti odprtokodne programske opreme, ker je spletni iskalnik v bistvu programska oprema. Ljudje se običajno odločijo za odprtokodne rešitve, ker je (Webb, 2001):

- odprtokodna programska oprema poceni za uporabo ali zastoj;
- odprtokodna programska oprema pogosto bolj prilagodljiva kot zaprtokodna (programska oprema bo brezhibno delovala na starejših računalnikih);
- odprtokodni proizvodi ponujajo prilagodljive funkcionalnosti in rešitve. Uporabnik tako lažje najde aplikacijo (takšen spletni iskalnik), ki ima točno tiste funkcionalnosti, kot jih potrebuje, ker razvijalci nenehno prispevajo nove rešitve v aplikacijo. Odprtokodna aplikacija je tudi veliko lažje prilagodljiva;
- ob priljubljenih odprtokodnih programih je ponavadi veliko razvijalcev, ki pomagajo ostalim uporabnikom preko forumov, elektronskih konferenc in elektronske pošte.

Slabosti odprtokodnih programskih rešitev pa so (Webb, 2001):

- podpora uporabnikom, ki sem jo navedel kot prednost odprtokodnih rešitev, je lahko tudi slabost. Ponavadi programske rešitve ne vsebujejo neke podrobnejše tehnične dokumentacije ali navodil uporabnikom. Prav tako je lahko sporazumevanje z drugimi uporabniki, ki prispevajo k programski opremi, težavno (jezik sporazumevanja, preveč tehnično izražanje ...);
- odprta koda ni proizvod »iz škatle«. Nalaganje in postopek, ko programsko opremo prilagajamo našim potrebam, je pogosto težaven in dolgotrajen. Mnogokrat je za

umestitev programske opreme na računalnik in za pravilno delovanje le-te potrebno programersko znanje.

## 2.2 Semantični iskalniki

Semantični iskalniki želijo poleg drugih kriterijev za razvrščanje rezultatov (popularnosti spletnih mest, ključnih besed ...) vnesene ključne besede predvsem razumeti, upoštevati širši pomen in kontekst, nato pa ponuditi najboljše rezultate (Dietze & Schröder, 2007; Iskold, 2006).

Primeri iskalnikov, ki delujejo na način procesiranja naravnega jezika (semantike), so: Hakia (hakia.com), Evri (evri.com), Powerset (powerset.com) (Radhakrishnan, 2009).

### Primeri semantičnega iskanja v spletnih iskalnikih

Primerjal sem iskalnik Hakia in iskalnik Najdi.si. Ker je Hakia spletni iskalnik, ki dobro deluje le v angleškem jeziku, sem ključno besedo oz. stavčno zvezo vnesel v angleškem jeziku, v Najdi.si pa v slovenskem.

*Tabela 1: Prikaz semantičnega iskanja*

	<b>Primer 1:</b>	<b>Primer 2:</b>
<b>Iskalnik / ključne besede (fraz)</b>		
Najdi.si	»Kako led vpliva na populacijo polarnih medvedov?« Ni rezultata	»populacija medved led« 1
Hakia	»How does ice affect polar bear populations?« 1	»population bear ice« 1

Tabela 1 prikazuje poskus iskanja informacije o vplivu ledu na populacijo polarnih medvedov. V prvem primeru sem oblikoval stavek, kot bi zastavil vprašanje. Prvi rezultat v iskalniku Hakia je bil odličen – uradna spletna stran Alaska Science Center, članek o raziskavah o polarnih medvedih v relaciji z ledom ([http://alaska.usgs.gov/science/biology/polar\\_bears/pbear\\_sea\\_ice.html](http://alaska.usgs.gov/science/biology/polar_bears/pbear_sea_ice.html)). Po drugi strani pa Najdi.si ni našel niti enega zadetka.

V drugem primeru sem iskal enako informacijo, vendar sem v vnosno polje vnesel le ključne besede, ki se nanašajo na moj interes. V iskalniku Hakia je bil prvi zadetek enak kot v prvem primeru, kar razumem kot dober rezultat. V iskalniku Najdi.si pa je prvi rezultat članek o ogroženosti severnega medveda na novičarski strani 24ur.com (<http://24ur.com/ekskluziv/zanimivosti/severni-medved-ogrozen.html>), drugi rezultat pa je članek o severnem medvedu v Wikipediji.

Najdi.si je geografsko lokalni iskalnik in ima zato v svojem indeksu mnogo manj spletne vsebine. Članki o polarnih medvedih so zato redki, kar je razlog, da nam ni bil ponujen boljši rezultat. Prva dva rezultata vseeno razumem kot dobra, saj o omenjeni temi v slovenskem jeziku verjetno ni boljše vsebine.

Lahke poizvedbe, kot je na primer iskanje po spletni strani bloga Jonasa Žnidaršiča, kjer v iskalnik vpišemo ključne besede »blog«, »jonas«, »znidarsic«, je mogoče rešiti s statističnimi algoritmi, kot jih ima večina klasičnih iskalnikov (na primer Google).

Malo težja poizvedba, kjer iskalnik sprašujemo na primer, »Kdaj je bil rojen Leonardo Da Vinci?«, je rešljiva le, če spletni iskalnik pozna nekaj zakonitosti naravnega jezika, v katerem uporabnik sprašuje.

Težja poizvedba, kjer iskalnik sprašujemo, »Kateri poslanec je nezakonito pridobil denar od davkoplačevalcev?«, potrebuje nekaj osebnih lastnosti in podatkov o uporabniku, računalnik, ki poda neko rešitev, in bazo podatkov, iz katere črpa določene podatke. Rezultat iskanja bi moral ponuditi en odgovor, ker je tudi uporabnik spraševal po enem poslancu.

Takšnim poizvedbam lahko pomaga umetna inteligenca oziroma semantična tehnologija. Semantična tehnologija naj bi reševala predvsem zapletene poizvedbe. Semantična tehnologija se lahko izkaže za zelo učinkovito pri problemih, ki jih rešujemo z relacijskimi bazami, torej pri poizvedbah po svetovnem spletu, kjer so potrebni podatki iz relacijskih baz. Svetovni splet pravzaprav je ogromna relacijska baza (Iskold, 2008).

## **2.3 Socialni iskalniki**

Socialno iskanje oziroma socialni iskalnik je spletni iskalnik, pri katerem na rezultate iskanj vplivajo oziroma jih presojujejo ljudje. Presojanje ljudi pomeni, da nekdo ali navadno kar množica ljudi presoja o tem, ali je neka vsebina vredna tega, da jo priporoči drugim uporabnikom iskalnika. Socialni iskalnik presoja spletne strani na podlagi uporabnikov,



običajen spletni iskalnik pa to opravi s pomočjo matematičnih in statističnih algoritmov (Sherman, 2006; Roush, 2007).

Prvi takšni iskalniki so v svojem bistvu spletni direktoriji, kjer so uredniki presojali, ali je določena spletna stran vredna vpisa v direktorij. V sodobnejšem času se vpisovanja v direktorij, iz katerega potem iskalnik išče rezultate, lahko poslužuje množica ljudi (angl. *crowdsourcing*). Socialni ali človeško gnani iskalniki lahko delujejo tudi na način, kjer nekdo ali množica ljudi razvršča rezultate, ki jih sprva prikaže iskalni algoritem.

Primeri socialnih iskalnikov so: Mahalo (mahalo.com), Wikianswers (answers.wikia.com/wiki/Wikianswers), Qitera (qitera.com), Stumpedia (stumpedia.com), ChaCha (chacha.com).

Socialni iskalniki se zaradi precej širokega pojma »socialnost« v literaturi pojavljajo kot iskalniki, ki delujejo na različne načine, zato jih je smiselno razdeliti na naslednje podkategorije:

- socialne iskalnike,
- iskalnike s povratno zvezo o relevantnosti (angl. *relevance feedback*),
- človeško gnane iskalnike.

### **Socialni iskalniki**

Socialni iskalniki so iskalniki, kjer na rezultate iskanj vpliva socialna skupnost uporabnika ali socialna skupnost, za katero iskalnik meni, da je sorodna uporabniku (Sherrets, 2007; Sherrets, 2008).

Primer takšnih iskalnikov so iskalniki v spletnih trgovinah, kjer spletna trgovina prikaže izdelke, ki so jih kupovali drugi ljudje ali pa jih je uporabnik že pregledal, vendar se jih ni odločil kupiti. Izbira teh izdelkov je lahko na podlagi demografskih lastnosti uporabnika, njegovega klikanja po določenih izdelkih in mnogih drugih karakteristik (Scime, 2004, str. 30-31).

Prednost socialnih iskalnikov je personalizacija rezultatov iskanj. Takšen spletni iskalnik lahko iz socialne mreže uporabnikovih prijateljev ali demografske in kulturne sorodnosti z uporabnikom pridobi informacije, ki pripomorejo k boljšim rezultatom iskanja (Arrington, 2008).

## **Iskalniki s povratno zvezo o relevantnosti**

V svojem diplomskem delu za človeško gnane iskalnike imenujem tiste iskalnike, kjer lahko nek uporabnik, ki je označen kot zanesljiv, iz množice uporabnikov odda svojo povratno informacijo neki iskalni poizvedbi na odprt, transparenten in javen način. Takšen uporabnik lahko za nek rezultat glasuje, doda novega ali izbriše obstoječega. Takšni iskalniki delujejo po principu uporabe množice (angl. *crowdsourcing*) (Sproose, 2009).

Crowdsourcing je pojem, ki pomeni zunanje izvajanje nekega opravila neke skupine posameznikov ali delno organizirane skupine posameznikov (množice) (Wiktionary, 2009).

Uporaba množice je deležna številnih kritik, po drugi strani pa obstajajo tudi uspešni primeri. Ustanovitelj Wikipedije, Jimmy Wales, trdi, da je za uspešno uporabo množice treba zgraditi uspešno strukturo, ki jo bodo nato uporabniki lahko uporabljali in v njej sodelovali na uspešen način. Ne sme pa pustiti vsega dela množici (McNichol, 2009).

Slabost človeško gnanih iskalnikov je možnost subjektivnih rezultatov, čeprav se je na primeru Wikipedije izkazalo, da zna množica ljudi zelo dobro ustvariti vsebino, ki je relevantna, zlorab pa je malo oziroma se hitro odkrijejo in odpravijo.

Slabost takšnih socialnih iskalnikov je, da iskalnik deluje dobro le takrat, ko je za neko ključno besedo že iskala množica ljudi pred nami, saj lahko le takrat uporabniki zadovoljivo razvrščajo rezultate in ponudijo najboljši rezultat. Izvršni direktor iskalnika Ask (ask.com) trdi, da na njihovem iskalniku vsak dan beležijo 60 % iskanj ali kombinacij ključnih besed, ki jih pred tem ni še nihče vpisal v iskalno poizvedbo (Liedtke, 2007).

Uspešna raba takšnih iskalnikov je pri iskanju po slikovni in video vsebini, kjer algoritmi in računalniki ne morejo vedeti, kaj je na sliki, saj ljudje bolje opišemo in opredelimo takšne dokumente kot matematični algoritmi.

## **Človeško gnani iskalniki**

Človeško gnani spletni iskalniki so spletni iskalniki, ki namesto matematičnih algoritmov za iskanje uporabijo ljudi. Cilj takšnih človeško gnanih spletnih iskalnikov je, da uporabniku kot rezultat prikažejo nekaj zelo relevantnih rezultatov.

Človeško gnani iskalnik ChaCha (chacha.com) išče informacije na način, da poizvedbo zastavi preko vnosne forme, rezultat pa namesto računalnika poišče človek, ki deluje na strani spletnega iskalnika (ChaCha, 2009).

Slabost takšnega iskalnika je, da iskalci informacij na strani spletnega iskalnika niso nujno najbolj usposobljeni ljudje. Čas, ki ga uporabnik porabi, da najde želeno informacijo, pa navadno ni krajši, kot če bi informacijo poiskal sam preko običajnih spletnih iskalnikov. Vprašanje, ki se mi zastavlja, je, kateri iskalnik uporabljajo iskalci, ki so zaposleni s strani iskalnika, ter ali imajo dovolj zaposlenih iskalcev v vsakem trenutku, ki lahko zadovoljijo vse uporabnike hkrati, in ali je čas, ko uporabnik vpiše iskalno poizvedbo v vnosno polje do prejema rezultata, dovolj kratek.

## 2.4 Vizualni iskalniki

Vizualni spletni iskalnik (angl. *visual search engine*) je spletni iskalnik, ki rezultate iskanj prikaže na interaktiven, grafičen način. Vizualni spletni iskalniki tudi omogočajo, da lahko sami urejamo in prilagajamo nadaljnja iskanja. Najbolj znani primeri vizualnih spletnih iskalnikov so Quintura (quintura.com), Kartoo (kartoo.com), Grokker (grokker.com) in Ujiko (ujiko.com).

Kartoo je meta iskalnik, kar pomeni, da nima svojega lastnega iskalnega algoritma, temveč uporablja različne spletne iskalnike, katerih rezultate potem združi. Meta iskalnike podrobneje predstavim v naslednjem poglavju. Rezultati so prikazani v dveh delih. Levi stolpec je enak kot pri iskalnikih, ki delujejo po konceptu gručenja. Na desni strani pa je prikazan interaktivni zemljevid nekih spletnih strani s ključnimi besedami in povezavami med njimi. Ob prehodu miške čez določeno ikono spletne strani ali ključne besede se pojavijo povezave izbrane spletne strani z drugimi. Te povezave naj bi imele semantično vrednost, saj nam iskalnik na ta način skuša pomagati pri iskanju sorodnih tem. Če kliknemo na določeno ključno besedo, se naš iskalni izbor zoži (ponudijo se novi rezultati), na ta način nam iskalnik na enostaven način pomaga hitreje najti iskano informacijo ali spletno stran. Prednost iskalnika Kartoo je predvsem enostavnost uporabe, slabost pa, da rezultati niso preveč zadovoljivi (Boswell, 2009; Kartoo, 2009).

Drugi primer vizualnega iskalnika je Quintera. Tudi Quintera prikazuje informacije na vizualen način in ne le kot besedilo (seznam). Quintera v nasprotju z iskalnikom Kartoo, ki prikazuje ikone spletnih strani ter ključnih besed, le-te prikazuje v obliki značk (angl. *tag*). Poleg tega pa na desni strani prikazuje rezultate v obliki seznama. Takšen iskalnik bi zato lahko poimenovali tudi spletni iskalnik značk (angl. *Tag Cloud Search Engine*). Do želenega

rezultata nam pomaga oblak značk (angl. *tag cloud*), saj nam iskalnik prikaže značke, ki so povezane s ključnimi besedami, ki smo jih vpisali na začetku. Ob prehodu miške čez določeno značko, se rezultati spremenijo (izbrana značka je dodana k ključnim besedam prvotne poizvedbe). Značke v oblaku značk pa lahko tudi odstranimo in tako iskalniku povemo, da nas ne zanimajo.

Glavna prednost iskalnika Quintera je enostavnost uporabe. V oblaku značk vidimo termine, na katere se prej sami nismo spomnili. Na ta način lahko najdemo informacije, ki jih običajno ne bi našli (Boswell, 2009).

V svojem bistvu so vizualni iskalniki enaki kot običajni iskalniki, kot jih poznamo (Google, Najdi.si), spremenjen je le uporabniški vmesnik z rezultati. Rezultati niso prikazani kot seznam, kjer je na prvem mestu najbolj relevanten rezultat, niže pa vedno manj relevantni. Takšni iskalniki lahko postanejo bolj priljubljeni, ko bodo na svoj vizualni način prikazali informacije ali podatke, ki jih ne moremo videti s seznama. Primer so semantične povezave, ki jih prikazuje iskalnik Kartoo.

## **Gručenje**

Gručenje (angl. *clustering*) pomeni dodeljevanje več objektov v gruče. Spletni iskalniki, ki uporabljajo koncept gručenja, rezultate ponudijo v gručah. Gruča je skupina podobnih tem, ki so relevantne iskani poizvedbi. Spletni iskalnik na ta način usmerimo v temo, ki nas zanima. Na ta način dobimo več bolj podrobnih rezultatov na izbrano temo. Primer takšnega iskanja prikazujem s spletnim iskalnikom Clusty (clusty.com), kjer poizvedujem po bananah. Poizvedbo sem naredil v angleščini (angl. *bananas*), ker iskalnik ne zna iskati po slovenskem spletu.

Slika 1: Primer iskanja z iskalnikom Clusty

The screenshot shows the Clusty search engine interface. At the top, there is a navigation bar with links for 'web', 'news', 'images', 'wikipedia', 'blogs', 'jobs', and 'more'. A search bar contains the word 'bananas' and a 'Search' button. Below the search bar, there are tabs for 'clusters', 'sources', and 'sites'. The main content area displays search results for 'bananas'. On the left side, there is a sidebar with a list of categories: 'All Results (219)', 'Plants (25)', 'Pictures (17)', 'Children (18)', 'Eating (16)', 'Music (14)', 'Facts (11)', 'History Of Bananas (7)', 'Diseases (7)', 'Favorite, Fruit (7)', and 'Comedy (6)'. The main results list shows three items:

- Banana - Wikipedia, the free encyclopedia**  
Banana is the common name for a fruit and also the herbaceous plants of the genus Musa which produce this commonly eaten fruit. They are native to the tropical region of Southeast Asia. Bananas are likely to have been first domesticated in Papua New Guinea. Today, they are cultivated throughout the tropics. Banana plants are of the family Musaceae. Botany · Properties · Trade · History · Cultivation · Pests, diseases and ...  
[en.wikipedia.org/wiki/Banana](http://en.wikipedia.org/wiki/Banana) - [cache] - Live
- Banana Plants - Musa Bananas For Growing Outdoors Or Indoors - Easy**  
Banana Plants - Musa Banana Plants and Trees producing Bananas ... Alocasia, Araucaria Monkey Puzzle Tree, Arbutus Strawberry Tree, Bamboos, Bananas, Bismarckia Palm, Butia Palm Tree.  
[www.easytropicals.com/pcp/Bananas.html](http://www.easytropicals.com/pcp/Bananas.html) - [cache] - Ask
- Ho-Ti Nursery**  
A plant nursery in Hawaii specializing in hardy tropicals including; Hawaiian ethnobotanicals, Anthuriums, Bananas, Amazonian sacred plants and medicinals.  
[www.ho-tinursery.com](http://www.ho-tinursery.com) - [cache] - Open Directory

Na levi strani se pojavijo gruče ali skupine, kjer s klikom izberem tisto, ki sem jo iskal. Rezultati se nato omejijo le na izbrano gručo.

Prednost koncepta gručenja je predvsem pri iskanjih, kjer uporabnik išče več informacij o točno določeni niši ali gruči. Področja, kjer je takšen iskalnik uporaben, so biologija, medicina in raziskovanje trga, kjer lahko trg ločimo na različne segmente (gruče) in raziskujemo njihove medsebojne povezave.

Zanimivi primeri iskalnikov poleg že omenjenega Clustyja, ki delujejo po konceptu gručenja, so še Quintera (quintera.com), Kartoo (kartoo.com) in Mnemomap (mnemo.org).

## Iskalniki s predogledi

Primer spletnih iskalnikov, ki skušajo uporabniku olajšati izkušnjo z uporabniškim vmesnikom, so spletni iskalniki, ki omogočajo predoglede. Uporabniku prikažejo del vsebine informacij, še preden uporabnik klikne na povezavo.

Iskalniki s predogledi skušajo skrajšati čas, ko uporabnik najde informacijo, ki jo je iskal. Primera takšnih spletnih iskalnikov sta Snap (snap.com) in Live Image Search (live.com/?scope=images).

## 2.5 Meta iskalniki

Meta iskalniki so iskalniki, ki uporabnikov zahtevke pošljejo na več drugih spletnih iskalnikov ali/in podatkovnih baz, rezultate pa nato prikažejo v agregatni obliki.

Namen in vizija meta iskalnikov sta, da je svetovni splet prevelik, da bi ga lahko v svoj indeks shranil le en spletni iskalnik. Zato naj bi dobili bolj celostne in popolne rezultate s kombiniranjem rezultatov, pridobljenih iz več iskalnikov. Meta iskalniki naj bi uporabnikom tudi olajšali iskanje preko več različnih iskalnikov.

Glavni predstavniki meta iskalnikov so Metacrawler (metacrawler.com), Webcrawler (webcrawler.com), Dogpile (dogpile.com) in Info (info.com).

Glavna prednost meta iskalnikov je iskanje po več indeksih hkrati, zato je podatkov in informacij, po katerih iščejo, veliko več. Po drugi strani pa rezultatov ne znajo prikazati bolj relevantno, saj ne poznajo internih povezav v bazah podatkov spletnih iskalnikov, ki jih vključujejo v iskanje. Pogosto prikažejo preveč rezultatov, ki niso točni. Glavna pomanjkljivost meta iskalnikov je zato ta, da ne znajo dobro razporediti različnih rezultatov, ki jih pridobijo iz ostalih spletnih iskalnikov.

Z nadaljnjim razvojem spletnih iskalnikov se bom ukvarjal v naslednjih poglavjih, kjer bom podrobneje analiziral, katere težave skušajo reševati določeni koncepti. Analiziral bom tudi, katere rešitve ponuja glavni igralec (Google) ter kaj ponuja glavni igralec na slovenskem trgu – Najdi.si.

## 3 TRŽNI VODJE

Za analizo stanja sem analiziral tržne vodje med spletnimi iskalniki. Danes je najbolj priljubljen spletni iskalnik Google z največ uporabniki in največ iskanji.

Tabela 2: Iskalniki glede na število iskanj

<b>Svetovni spletni iskalniki</b>		
	Iskanja (1000 mio)	Tržni delež
Skupaj internet	66,221	100,0
Googlove strani	41,345	62,4
Yahoojeve strani	8,505	12,8
Baidu Inc.	3,428	5,2
Microsoftove strani	1,940	2,9
NHN Corporation	1,572	2,4
eBay	1,428	2,2
Time Warner Network	1,062	1,6
Ask Network	728	1,1
Yandex	566	0,9
Alibaba.com Corporation	531	0,8

Vir: ComScore, 2008.

### 3.1 Google

*Računalnik v filmu Star Trek me ne navdušuje. Zastavljajo mu neke naključne poizvedbe, računalnik pa dolgo časa razmišlja o rezultatu poizvedbe. Mislim, da mi to lahko storimo bolje (Larry Page, 2004).*

Google je horizontalni iskalnik, saj podatke išče po širokem spektru informacij, po čim širšem področju svetovnega spleta. Google je trenutno najbolj popularen spletni iskalnik, ki se želi uporabnikom že od svojega obstoja prikupiti s hitrim delovanjem in kakovostnimi zadetki iskalnih poizvedb. Googlova tehnologija razvrščanja zadetkov velja za zelo varovano skrivnost. Nihče resnično ne pozna vseh kriterijev, po katerih Google razvršča rezultate, zato bom v diplomskem delu poskusil pojasniti nekaj kriterijev, ki vplivajo na zadetke.

Googlov koncept razvrščanja rezultatov temelji na podlagi popularnosti povezav (PageRank kriterija). Google PageRank opisuje kot demokratično načelo glasovanja spletnih strani s svojimi povezavami na druge strani. Uporablja strukturo povezav med spletnimi stranmi in njihovo težo. Google tolmači povezavo z strani A na stran B kot glas strani A strani B. Vendar Google ne obravnava le števila povezav (glasov), ki jih določena stran prejme, ampak analizira tudi težo glasu, ki ga prejme. Strani, ki so »pomembne«, dajo večjo težo s svojim glasom strani, ki ji dodelijo glas. Spletna stran, na katero kaže mnogo povezav z visokim PageRank kriterijem, tudi sama dobi visok PageRank. Po drugi strani glasovi iz »pomembnih« strani ne pomenijo nič, če se ne ujemajo z iskalnim kriterijem (iskalnim nizom). Google kombinira napredne tehnike ujemanja besedil s pomembnostjo strani z glasovi, ki jih prejme (Google, 2009).

PageRank, ki je Googlu pomagal postati najbolj priljubljen spletni iskalnik, pa vedno bolj kaže svoje slabe lastnosti, saj rezultati iskanj niso vedno najboljši. Težava tehnologije PageRank je ta, da ustvarjalci spletnih strani – spletni skrbniki izrabljajo šibko točko povezovanja in »umetno« dodelijo povezave (glasove) na spletno stran, za katero želijo, da se pojavlja više na rezultatnih straneh spletnega iskalnika. Google mora zato upoštevati še več dejavnikov spletnih strani in zato zmanjšuje pomen PageRank tehnologije.

Google ima poleg PageRank kriterija še preko 150 drugih kriterijev, po katerih razvršča rezultate.

Google želi ostati vodilni spletni iskalnik in zato tudi sam razvija nove koncepte in tehnologije, ki izboljšujejo iskanje podatkov. Pri svojih iskanjih uporablja umetno inteligenco



(semantiko), razvija vertikalne iskalnike, je socialen iskalnik, razvija tudi vedno nove načine prikazovanja rezultatov (CNN Money, 2009).

Google uporabniku omogoča, da označi spletne strani, ki jih ima za bolj relevantne in jim zaupa. Uporabnik lahko označi lokalne spletne strani ali pa priljubljene novinarske strani. V rezultatih iskanj bodo te strani večkrat prikazane in označene kot prednostne strani (Cutts, 2009).

Google se veliko ukvarja z različnimi prikazi rezultatov. Rezultate zna prikazati kot časovni trak, na zemljevidu ali v kontekstu z ostalimi tipi informacij (npr. datumom, merami, lokacijo in sliko, matriko ...) (Google Labs, 2009).

Slabost Googla je premalo vključenih dokumentov v iskanje in iskanje po zaprtih bazah podatkov (npr. iskanje znotraj Facebooka, akademskih vsebin ...) ter strukturiranih podatkih. Strukturirani podatki so baze podatkov v globokem spletu (angl. *deep web*). Globoki splet pomeni vsebino, ki je za spletnimi obrazci in je ni mogoče indeksirati kot običajne spletne strani. Google zato vsebino trenutno pridobi tako, da avtomatično izpolnjuje spletne obrazce in pridobljene rezultate shranjuje v svoj indeks (Perez, 2009).

Google v svojem indeksu nima shranjenih vseh strani (npr. nepovezanih strani, do katerih iskalni črv ni prišel, namenoma izločenih strani, baz podatkov ...).

### **Personalizirano iskanje**

Personalizacija je način, da proizvod prilagodimo uporabniku po meri. Personalizacija rezultatov iskanja preko spletnega iskalnika zato ponudi rezultate, ki so prilagojeni uporabniku. Personalizacija vključuje tudi lokalizacijo oz. regionalizacijo. Regionalizacija je torej enostavnejša oblika personalizacije, saj personalizacija upošteva še dodatne kriterije (Levene, 2006, str. 128-129).

Po mnenju Matta Cuttsa ima personalizacija trenutno največje možnosti, da resnično uspe izboljšati iskalne rezultate. Tako uspešna pa je lahko zato, ker se mora uporabnik le odreči delčku zasebnosti, v zameno pa iskalnik postori vse ostalo in postreže z bistveno boljšimi rezultati, ki so popolnoma brezplačni (MacManus, 2008).

Primer: če v Sloveniji iščemo po ključni besedi »banka«, bomo dobili rezultate slovenske banke, v Nemčiji pa nemške banke (Schiffman, 2008).

Če želi spletni iskalnik ponuditi personalizirane rezultate, jih prilagodi glede na:

- zgodovino iskanj (kakšne vsebine je uporabnik večkrat iskal v preteklosti);
- shranjene priljubljene povezave (katere povezave ima shranjene med svojimi priljubljenimi povezavami – imeti mora nameščeno dodatno orodno vrstico v svojem brskalniku ali jih ima shranjene v svojem iGoogle računu);
- način (vedenje) uporabe v različnih spletnih skupnosti (npr. FaceBook, YouTube);
- izbiro rezultatov med ponujenimi;
- čas, ki ga uporabnik porabi na določeni spletni strani, ki jo je izbral med rezultati.

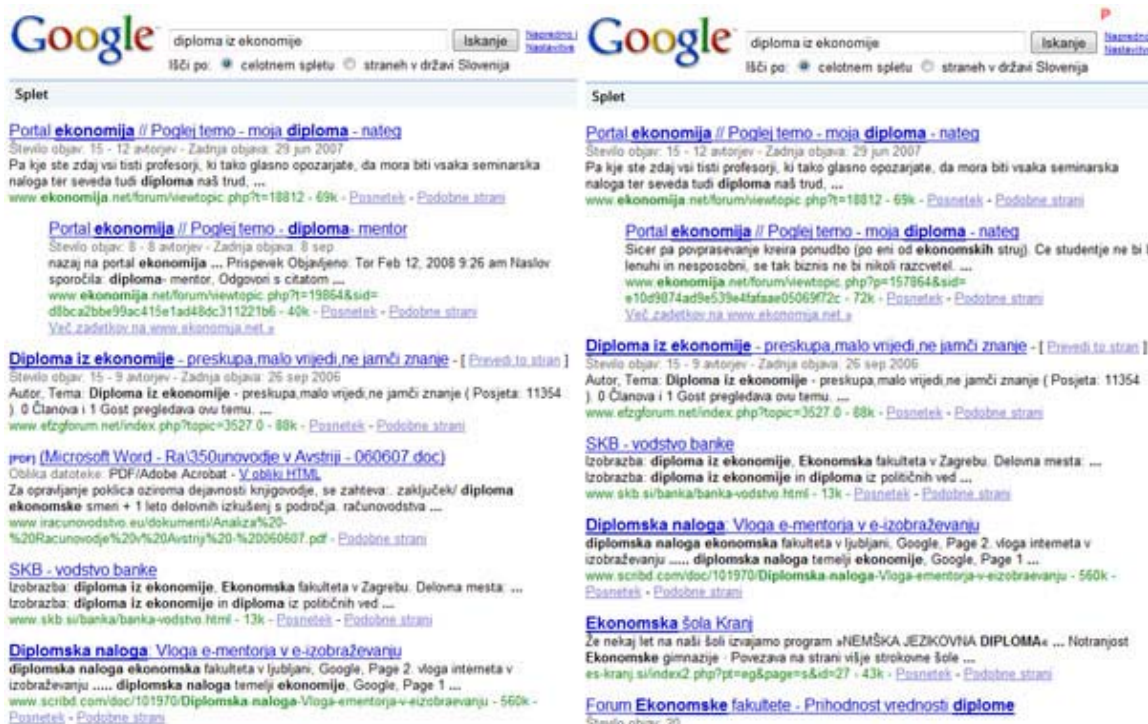
Načini, kako lahko uporabnik prilagodi oziroma personalizira rezultate, so različni. Med bolj pogostimi načini so (Google Experimental, 2009):

- dodaj med zaznamke (angl. *bookmarks*),
- označevalec ali značka (angl. *tagging*),
- glasovanje (angl. *voting*).

Spletni iskalnik lahko nato na podlagi množice podobnih ali enakih glasovanj, označevanj ali zaznamkov uporabnikov svoje rezultate popravi in tako pri vseh naslednjih iskanjih prikaže boljše rezultate. Personalizacijo rezultatov Google ponudi le uporabnikom, ki imajo svoj Google račun (in so v njega tudi prijavljeni). Podatke o uporabi iskalnika pa črpa iz Googlovega računa (Sullivan, 2007).

Med najbolj znanimi spletnimi stranmi, ki so uporabljale personalizirane rezultate, je spletna trgovina Amazon (amazon.com). Uporabnikom priporoča izdelke med rezultati ali na prvi strani, za katere misli, da so uporabniku najbližje (in jih želi kupiti). Določene izdelke lahko ponuja glede na uporabnikove prejšnje nakupe v spletni trgovini.

Slika 2: Personalizirani in nepersonalizirani rezultati v Googlu



Vir: Google, 2009.

Primer personaliziranih rezultatov prikazujem na zgornji sliki 2, kjer je na levi prikazano običajno, nepersonalizirano iskanje, na desni pa personalizirano iskanje, kjer je uporabnik prijavljen v Googlov račun. Opazim precejšnje razlike v rezultatih. Četrtega rezultata pri nepersonaliziranih rezultatih iskalnik ne prikazuje pri personaliziranih rezultatih, saj predvideva, da rezultat ni relevanten za moje iskanje. Enako stori z drugim rezultatom pri nepersonaliziranem iskanju, saj ga nadomesti z drugim rezultatom.

## Regionalizacija

Največji svetovni spletni iskalniki (Google, Yahoo, MSN ...) so prilagojeni ameriškim uporabnikom, čeprav naj bi bili namenjeni vsem uporabnikom. Ti isti spletni iskalniki tudi nimajo posebnega iskalnika izključno za ameriški trg. Po drugi strani pa so regionalizirani iskalniki namenjeni določeni geografski ali jezikovni regiji.

## **Uporabniški vmesnik**

Uporabniški vmesnik je v regionalizirani različici narejen v jeziku, ki mu je namenjen. Da je uporabniški vmesnik oziroma spletni iskalnik kot spletna stran v določenem jeziku, pa naj ne bi vplivalo na iskalne rezultate.

## **Izključevanje domen**

Pri filtraciji domen bo spletni iskalnik v svoj indeks vključil le spletne strani, ki so namenjene določeni državi. Primeri domenskih končnic, namenjenih določeni državi:

- Slovenija – .si,
- Hrvaška – .hr,
- Velika Britanija – .co.uk.

Pri domenskem izključevanju obstajata dve veliki težavi. Prva težava je ta, da ima mnogo podjetji registrirane domene, ki se končajo s .com ali .net. Te spletne strani morajo biti zato ročno vnesene v indeks.

Druga težava je ta, da večina držav nima izključno svojih končnic, saj jih ponujajo tudi kupcem.

Težavo v prihodnosti lahko povzročijo tudi nove domenske končnice, ki so namenjene splošni uporabi (npr. .web). Te domenske končnice so popolnoma neodvisne od države.

## **Rezultati glede na jezik**

Večina spletnih iskalnikov danes prepozna jezik na spletnih straneh in jim tako lahko določi regijo. Spletni iskalniki vidijo, v katerem jeziku je spletna stran iz meta podatkov ali pa po določenih besedah, ki jih iskalnik prepozna, da pripadajo določenemu jeziku.

## 3.2 Najdi.si

Tako kot Google lahko tudi Najdi.si uvrščamo med horizontalne iskalnike. Značilnost Najdi.si je, da je geografsko omejen spletni iskalnik, kar pomeni, da podatke išče po geografsko omejenem področju – slovenskem spletu.

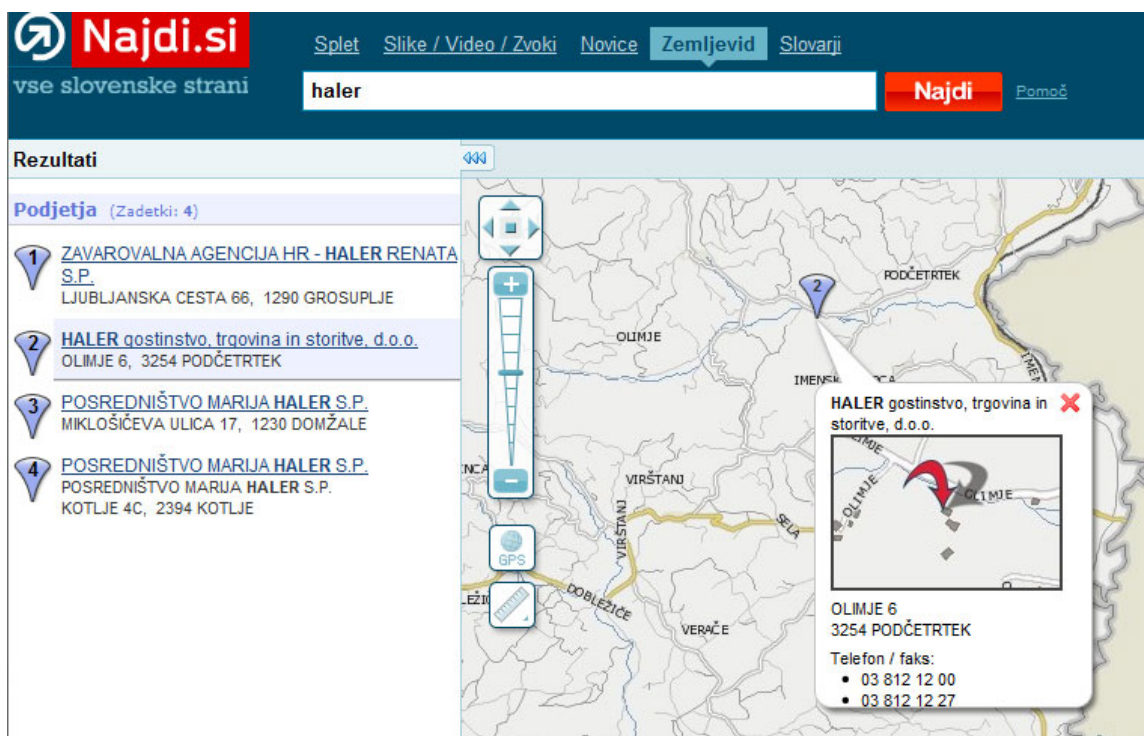
Spletni pajki v indeks Najdi.si vpisujejo spletne strani, ki izpolnjujejo enega od dveh pogojev:

- imajo domensko končnico .si, tudi če stran ne gostuje na strežniku v Sloveniji in tudi če vsebina ni v slovenščini,
- gostujejo na strežniku v Sloveniji.

Spletne strani, ki so na domenskih končnicah, drugačnih od .si, in ne gostujejo na strežniku v Sloveniji, vendar imajo slovensko vsebino, v indeks niso samodejno vpisane. Te strani je treba vpisati ročno, najpogosteje jih vpiše lastnik strani, vpis pa pregleda urednik. Najdi.si ni semantično podprt iskalnik in tako ne prepozna jezika, v katerem je neka spletna stran napisana. Očitno pa je, da so jezikovne posebnosti slovenskega jezika vgrajene v iskalnik, saj nam iskalnik ponudi sorodne besedne zveze, išče po dvojini in množini, čeprav smo ključno besedo vpisali v ednini, korenu besede pa zna dodajati smiselne končnice, kar izboljša iskalne rezultate.

Najdi.si vključuje tudi različna uporabna in manj uporabna orodja. Takšna orodja so na primer zemljevid, novice, avtomobili, imenik podjetij, vici, horoskop, vreme ... Prednost nekaterih orodij pred konkurenco je predvsem v lokalnosti. Primer je modul »vreme«, kjer si lahko uporabnik nastavi svojo lokacijo v Sloveniji. Prav tako je zemljevid Slovenije bolj natančen od konkurenčnih, saj ima shranjene naslove podjetij, ulice in kraje (vasi).

Slika 3: Primer iskanja z zemljevidom na Najdi.si



Vir: Najdi.si, 2009.

Pri iskanju po ključni besedi »haler« Najdi.si prikaže štiri rezultate. Med njimi je tudi pivnica, ki sem jo iskal. Po kliku na povezavo »haler gostinstvo, trgovina in storitve, d. o. o.« se prikaže podroben zemljevid z osnovnimi podatki o podjetju (naslov, kontakt).

Najdi.si želi biti slovenski vstop v internet in zato nekakšen portal informacij, ki zanima najširši krog ljudi. Najdi.si se v marsičem spogleduje z Yahoojem. Njegovo tehnologijo uporablja tudi pri iskanju po tujih spletnih straneh.

Prednost Najdi.si pred konkurenčnimi iskalniki je, da Najdi.si pozna jezikovne posebnosti slovenščine in je zato bolj uspešen pri določenih iskanjih. Najdi.si je zato bolj lokalni od iskalnika Google.si – iskalnika Google, namenjenega slovenskemu spletu.

Največja slabost spletnega iskalnika Najdi.si je ta, da v svojem indeksu nima shranjenih vseh spletnih strani. Google.si ima v svojem indeksu shranjenih mnogo več spletnih strani. Primer, ki potrjuje mojo tezo, je preprosto iskanje po poljubni temi. Če v oba iskalnika vpišem besedno zvezo »motorna vozila«, Najdi.si informacije išče po 1.458.918 spletnih straneh, Google.si pa po približno 4.020.000 (na dan 19. 5. 2009).

### 3.3 Izzivalci

Spletni iskalniki, ki so trenutno najbolj napredni in najbolj navdušujejo strokovnjake spleta, so iskalniki, ki skušajo tržne vodje premagati na področjih, kot so izboljšanje tehnologije relevantnosti rezultatov, izboljšanje prikaza rezultatov, ali pa želijo boljšo tehnologijo relevantnosti rezultatov in boljši prikaz rezultatov, vendar na ožjem področju iskanja. Težave skušajo reševati z naslednjimi koncepti:

- uporaba umetne inteligence oziroma semantike, s katero skušajo izboljšati relevantnost rezultatov in olajšati zahtevnost uporabe spletnega iskalnika;
- prikaz rezultatov na inovativen način, s čimer skušajo bolje prikazati iskane informacije;
- iskano poizvedbo prikažejo kot odgovor na vprašanje, s čimer želijo prikazati točne rezultate, vendar običajno le za bolj omejene vsebine.

Primeri iskalnikov, ki trenutno pridobivajo na priljubljenosti, so True Knowledge ([trueknowledge.com](http://trueknowledge.com)), Wolfram Alpha ([wolframalpha.com](http://wolframalpha.com)), Hakia ([hakia.com](http://hakia.com)) in PowerSet ([powerset.com](http://powerset.com)) (Sutter, 2009).

Wolfram Alpha je iskalnik, ki rezultat prikaže kot odgovor namesto spiska dokumentov ali spletnih strani. Wolfram Alpha prikazuje rezultate, ki sploh ne obstajajo na svetovnem spletu, dokler jih sami ne poiščemo. V tem smislu sploh ni spletni iskalnik, temveč ogromen računalnik z ogromno znanja (Dziuba, 2009; Lardinois, 2009).

Za iskalno poizvedbo po besedni zvezi »Munich to New York City«, ki jo prikazujem v prilogi 3, spletni iskalnik vrne mnogo podatkov, ki naj bi nas zanimali o letu iz Münchna v New York, npr. razdaljo leta, zemljevid poti, lokalni čas ... Rezultat, ki ga je iskalnik prikazal, morda ni bil zapišan nikjer na svetovnem spletu, na nobeni spletni strani, od koder bi klasičen spletni iskalnik pridobil rezultat. Rezultat, ki ga je podal Wolfram Alpha, je bil izračunan po kliku na gumb »compute«.

Po ustanoviteljevem mnenju niti ne gre za spletni iskalnik, temveč »računalniški iskalnik znanja« (angl. *Computational knowledge engine*), saj WolframAlpha po drugi strani ne zna odgovoriti na splošna vprašanja, ki jih uporabniki iščejo vsakodnevno. Iskalnik na primer ne pozna odgovora na vprašanje, kam naj gre uporabnik na nedeljski izlet.

Tudi socialni portali kot na primer FaceBook (facebook.com) razvijajo svoj spletni iskalnik. Čeprav iščejo le po svoji notranji bazi, na določene poizvedbe prikazujejo zelo dobre rezultate. To lahko storijo zato, ker poznajo socialne navade uporabnika in uporabnikovih prijateljev.

Spet druge socialne strani, kot je na primer Twitter (twitter.com), omogočajo iskanje po kratkih sporočilih uporabnikov. Uporabniku takšen spletni iskalnik omogoča poiskati aktualen podatek ali informacijo, torej omogoča pridobitev informacij v realnem času od nastanka nekega dogodka.

## **4 PRIHODNOST ISKALNIKOV**

Svetovni splet prihodnosti lahko poimenujemo Web 3.0, če je sedanja stopnja razvoja svetovnega spleta poimenovana Web 2.0. Termin Web 3.0 si vsak predstavlja nekoliko drugače, saj ne obstaja jasna slika prihodnosti. Mnogi spletni guruji pa imajo enoten pogled na nekatere tehnologije ali možne smeri razvoja, zato bom le-te podrobneje opisal. Spletni guru Nova Spivack Web 3.0 definira kot združitev različnih tehnoloških trendov, ki naj bi dosegli novo raven razvoja (Spivack, 2007).

Tehnološki trendi razvoja so:

- transformacija svetovnega spleta (vzpostavitev enotnega računalniškega okolja, v katerem bodo združene različne aplikacije, ki bodo programsko in procesorsko moč pridobile iz oddaljenih računalnikov);
- vseprisotna povezljivost (širokopasovni internetni dostop od koder koli s pomočjo različnih naprav);
- omrežno računalništvo;
- odprte tehnologije (vmesnik za programiranje, odprta koda, prosto dostopni podatki, prosti protokoli);
- odprta identiteta (OpenID in odprta identiteta preko socialnih spletnih portalov);
- inteligenten splet (semantičen splet, aplikacije ki so uporabniku bolj prijazne);
- porazdeljene baze podatkov (baza podatkov svetovnega spleta);
- inteligentne aplikacije (aplikacije, ki so uporabniku prijazne, saj jih lahko uporablja v svojem naravnem jeziku, samoučenje aplikacij ...).

Umetna inteligenca je termin, ki navdušuje največ ljudi pri spletnem razvoju in največ ljudi tudi verjame, da bo umetna inteligenca naslednji večji preskok v razvoju spleta in spletnih iskalnikov. Za umetno inteligenco se zato uporablja tudi termin Web 3.0.



Inteligentni splet predstavlja semantičen splet, iskanje v naravnem jeziku, podatkovno rudarjenje, strojno učenje in umetno inteligenco (Markoff, 2006).

Prihodnost spletnih iskalnikov ni jasna in težko bi napovedali, katera tehnologija nam bo služila v prihodnosti. V današnjem času se razvija množica spletnih iskalnikov, ki temeljijo na različnih konceptih. Koncepti, ki bi v prihodnosti lahko bili bolj aktualni, so iskalniki na podlagi (MacManus, 2009; Knight, 2009):

- umetne inteligence (primer: Hakia, Powerset),
- človeškega dela (človeško gnani iskalniki – primer: ChaCha),
- navpičnega iskanja (primer: SimplyHired, Technorati),
- personalizacije (primer: Collarity),
- gručenja (primer: Clusty, SearchMash),
- socialnega iskanja (primer: Eurekster, Rollyo),
- vizualizacije (primer: Quintura, Kartoo),
- iskanja s predogledi (primer: Snap, Live Image Search).

Spletni iskalniki vsakodnevno izpopolnjujejo svojo tehnologijo na področju iskalnih rezultatov, prikaza rezultatov in načina komunikacije z uporabnikom. Cilj spletnih iskalnikov je pridobiti čim več uporabnikov, ki jih lahko na različne načine »unovčijo« in z njihovo pomočjo služijo. Vizija spletnih iskalnikov pa je imeti zadovoljne uporabnike, ki bodo v najkrajšem času prišli do zelene informacije. Vizija in cilj se zato na nek način izločujeta, saj je zadovoljenega uporabnika težje unovčiti.

Spletni iskalnik prihodnosti vidim kot zmes konceptov in tehnologij, ki se prilagaja vsakemu uporabniku posebej. Spletni iskalnik, ki bo želel pridobiti znaten delež med uporabniki svetovnega spleta, bo moral združiti točnost in relevantnost rezultatov z enostavnostjo uporabe. Novi spletni iskalniki, ki se dnevno pojavljajo na svetovnem spletu, težijo k hitrejšim, pametnejšim, bolj osebnim in vizualno bolj relevantnim rezultatom (Segev, 2008; Bradley, 2006; Gord, 2007).

Za pridobitev opaznejšega tržnega deleža bo v prihodnosti treba narediti še kaj več, kot biti malenkost boljši od vodilnih spletnih iskalnikov. Mnogi novi iskalniki, na primer Cuil (cuil.com), prejmejo veliko odobravanja in zanimanja v začetku obratovanja, kasneje pa to občudovanje izgine. Google je svojo popularnost gradil v času, ko je svetovni svet rasel zelo hitro, svoje uporabnike je tako sam naučil, kako naj ga uporabljajo. Ljudje neradi spreminjamo svoje navade, po drugi strani pa nam Google še vedno v večini primerov postreže z dobrimi rezultati.

Google še naprej ostaja največji spletni iskalnik s približno 64 % vseh spletnih iskanj v Združenih državah Amerike (comScore, 2009).

Googlova dobra lastnost je, da ne počiva na lovorikah. V lanskem letu je podjetje samo na področju iskalnika naredilo 359 izboljšav. Nekatere izboljšave, kot na primer personaliziranje rezultatov, so dobro vidne, druge pa spet ne. Sergey Brin, ustanovitelj Googla, pravi, da ima Google vgrajene več inteligence, kot jo pokaže na prvi pogled, med drugim tudi umetno inteligenco, vendar le-ta še vedno ni na takšnem nivoju, kot bi si želeli zaradi zelo hitrega širjenja svetovnega spleta, kjer za vsako iskalno poizvedbo lahko dobimo množico dokumentov, ki so lahko slike, videi, knjige, zemljevidi ... (Brin, 2009).

## **SKLEP**

V diplomskem delu sem predstavil spletne iskalnike, njihove tehnologije in razvoj. Pojasnil sem, kako je potekal razvoj iskalnikov do danes in katere tehnologije bodo iskalnikom pomagale, da bodo rezultati iskanj v prihodnosti boljši. Menim, da spletni iskalnik prihodnosti ne more biti omejen le z eno od tehnologij, ki sem jih opisal, temveč s kombinacijo le-teh in njihovimi izboljšavami. Umetna inteligenca lahko iskalnikom pomaga pri uporabniškem vmesniku, človeško gnani iskalniki pa iščejo po specifičnih nišah.

Spletni iskalnik prihodnosti bo vedel, kaj ima uporabnik rad, kakšne običaje ima, kakšne interese ima ali je imel, iz kakšne kulture izhaja, kje je bil, kaj je kupil, kakšen jezik govori, kje živi, s kom se družijo ... Prenosni telefoni so zelo osebni in vsebujejo zelo veliko informacij o uporabniku in njegovi socialni mreži. Z uporabo interneta preko prenosnih telefonov ima spletni iskalnik prihodnosti veliko možnosti, da postreže boljše rezultate, saj lahko uporabnika bolje pozna.

Spletni iskalnik prihodnosti bo uporabniški vmesnik in bo rezultate iskanj prikazal na najbolj naraven in intuitiven način. Spletni iskalnik prihodnosti bo razumel uporabnika in njegov naraven jezik. Uporabnik ne bo razmišljal, katere ključne besede in kako naj zastavi poizvedbo, da bi dobil želen rezultat. Spletni iskalnik bo torej prikaz rezultatov prilagodil iskalni poizvedbi in uporabniku samemu (MacManus, 2009).

Spletni iskalnik lahko ponudi le tako dober rezultat, kolikor dobra je vsebina v dokumentu oziroma v rezultatu. Na svetovnem spletu še vedno manjka določena vsebina. Knjižnice imajo zelo veliko literature, ki ni objavljena na svetovnem spletu. Nekatera podjetja že masovno

skenirajo knjige in jih objavljajo. Proces je počasen, vendar so rezultati iskanj tudi zaradi tega iz dneva v dan boljši.

Popolnega spletnega iskalnika ni in ga morda tudi nikoli ne bo, ker bi popoln iskalnik podal odgovor na naše vprašanje, še preden bi uporabnik vprašanje sploh postavil (Schmidt, 2009).

Informacijska tehnologija se razvija zelo hitro, zato je zelo težko napovedovati njeno prihodnost.

*V letu 1981 se je povečanje notranjega spomina s 64 k na 640 k zdelo dovolj za deset let, vendar so uporabniki že po šestih letih ugotovili, da to predstavlja problem (Bill Gates, 1989).*

## SEZNAM LITERATURE IN VIROV

### Literatura

1. Asadi, S. & Jamali, H. R. (2004). Shifts in Search Engine Development: A Review of Past, Present and Future Trends in Research on Search Engines. Najdeno 26. aprila 2009 na spletnem naslovu <http://www.webology.ir/2004/v1n2/a6.html>.
2. Berry, M. W. & Browne M. (2005). *Understanding Search Engines: Mathematical Modeling and Text Retrieval (Software, Environments, Tools)*. Second Edition. Philadelphia PA: SIAM, Society for Industrial and Applied Mathematics; 2 edition.
3. Bhatia, M. P. S. & Khalid, A. K. (2008). Information retrieval and machine learning: Supporting technologies for web mining research and practice. Najdeno 27. aprila 2009 na spletnem naslovu <http://www.webology.ir/2008/v5n2/a55.html>.
4. Burnett, R. P. & Marshall, D. (2003). *Web Theory: An Introduction*. New York, NY: Routledge; 1st edition.
5. Dietze, H. & Schröder, M. (2007). GoWeb: A semantic search engine for the life science web. Najdeno 20. aprila 2009 na spletnem naslovu <http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-435/paper01.pdf>.
6. Levene, M. (2006). An introduction to search engines and Web navigation. Harlow, England: Addison Wesley Publishing Company; illustrated edition edition.
7. Scime, A. (2004), Web mining: applications and techniques. Hershey, PA: Idea Group Publishing; illustrated edition edition.
8. Segev, E. (2008). Search Engines and Power: A Politics of Online (Mis-) Information. Najdeno 7. aprila 2009 na spletnem naslovu <http://www.webology.ir/2008/v5n2/a54.html>.

## Viri

1. Arrington, M. (2008). Video Of Mahalo v. Wikia Search At DLD; Google's Marissa Mayer Weighs In. Najdeno 5. junija 2009 na spletnem naslovu <http://www.techcrunch.com/2008/01/21/mahalo-v-wikia-search/>.
2. Barker, J. (2003). What Makes a Search Engine Good? Najdeno 22. maj 2009 na spletnem naslovu <http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/SrchEngCriteria.pdf>.
3. Betsy Schiffman. (2008). Q&A With Google's Matt Cutts About SEO and the Future of Search. (2008). Najdeno 2. februarja 2009 na spletnem naslovu <http://blog.wired.com/business/2008/03/qa-with-googles.html>.
4. Blog Najdi.si. Najdeno 19. maja 2009 na spletnem naslovu <http://blog.najdi.si/>.
5. Boswell, W. A Tag Cloud Search Engine. Najdeno 12. aprila 2009 na spletnem naslovu <http://websearch.about.com/od/dailywebsearchtips/qt/dnt0830.htm>.
6. Boswell, W. Search with Kartoo, a Visual Meta Search Engine. Najdeno 12. aprila 2009 na spletnem naslovu <http://websearch.about.com/od/enginesanddirectories/a/kartoo.htm>.
7. Bradley, P. (2006). Search Engines: Where We Were, Are Now, and Will Ever Be. Najdeno 27. aprila 2009 na spletnem naslovu <http://www.ariadne.ac.uk/issue47/search-engines/>.
8. Brin, S. & Page, L. (2009). The Anatomy of a Large-Scale Hypertextual Web Search Engine. Najdeno 3. aprila 2009 na spletnem naslovu <http://infolab.stanford.edu/~backrub/google.html>.
9. Brin, S. (2009). Co-founder & President, Technology: The 2008 Founders' Letter. Najdeno 2. junija 2009 na spletnem naslovu <http://googleblog.blogspot.com/2009/05/2008-founders-letter.html>.
10. BusinessWeek. (2004). Google's Goal: "Understand Everything" – Larry Page intervju. Najdeno 12. aprila 2009 na spletnem naslovu [http://www.businessweek.com/magazine/content/04\\_18/b3881010\\_mz001.htm](http://www.businessweek.com/magazine/content/04_18/b3881010_mz001.htm).
11. ChaCha: How to ChaCha. Najdeno 5. junija 2009 na spletnem naslovu <http://answers.chacha.com/about-chacha/how-it-works/>.
12. CNN Money. (2009). One-on-one with Google's CEO. Najdeno 2. junija 2009 na spletnem naslovu <http://money.cnn.com/video/news/2009/06/01/news.google.schmidt.full.cnnmoney/>.
13. comScore. (2008). qSearch 2.0comScore qSearch 2.0, Worldwide Search Top 10. Najdeno 22. aprila 2009 na spletnem naslovu <http://www.comscore.com/press/release.asp?press=2018>.
14. comScore. (2009). comScore Releases April 2009 U.S. Search Engine Rankings. Najdeno 22. aprila 2009 na spletnem naslovu

- [http://www.comscore.com/Press\\_Events/Press\\_Releases/2009/5/comScore\\_Releases\\_April\\_2009\\_U.S.\\_Search\\_Engine\\_Rankings](http://www.comscore.com/Press_Events/Press_Releases/2009/5/comScore_Releases_April_2009_U.S._Search_Engine_Rankings).
15. Conntech Listserv Archives 2004: Yanoff's List - anyone remember? Najdeno 20. Januarja 2009 na spletnem naslovu  
<http://www.cslib.org/conntech/2004to05/0506.html>.
  16. Dziuba, T. (2009). Mathematica man brews 'AI' Google Killer. Najdeno 17. marca 2009 na spletnem naslovu [http://www.theregister.co.uk/2009/03/17/wolfram\\_alpha/](http://www.theregister.co.uk/2009/03/17/wolfram_alpha/).
  17. Enge, E. (2007). Search Engine Watch: Are Vertical Search Engines the Answer to Relevance? Najdeno 22. marca 2009 na spletnem naslovu  
<http://searchenginewatch.com/3624377>.
  18. Gates, B. (1989). Bill Gates talk on Microsoft. Najdeno 3. junij 2009 na spletnem naslovu  
<http://csclub.uwaterloo.ca/media/1989%20Bill%20Gates%20Talk%20on%20Microsoft.html>.
  19. Gerkeš, M. (2000). Iskalniki in iskanje na svetovnem spletu. Maribor: Institut informacijskih znanosti. Najdeno 7. aprila 2009 na spletnem naslovu  
[http://home.izum.si/cobiss/cobiss\\_obvestila/2000\\_3/Html/clanek\\_01.html](http://home.izum.si/cobiss/cobiss_obvestila/2000_3/Html/clanek_01.html).
  20. Google Experimental. Najdeno 4. februarja 2009 na spletnem naslovu  
<http://www.google.com/experimental/>.
  21. Google Labs. Najdeno 16. februarja 2009 na spletnem naslovu  
<http://www.googlelabs.com>.
  22. Google Webmaster guidelines. Najdeno 13. aprila 2009 na spletnem naslovu  
<http://www.google.com/support/webmasters/bin/answer.py?answer=35769>.
  23. Google Webmaster Guidelines. Najdeno 14. maja 2009 na spletnem naslovu  
<http://www.google.com/support/webmasters/bin/answer.py?answer=35769>.
  24. Google: Technology Overview. Najdeno 20. aprila 2009 na spletnem naslovu  
<http://www.google.com/corporate/tech.html>.
  25. Gord. Out of My Gord (2007, 2. marec). Matt Cutts Interview on Personalization and the Future of SEO. Najdeno 13. aprila 2009 na spletnem naslovu  
<http://www.outofmygord.com/archive/2007/03/02/Matt-Cutts-Interview-on-Personalization-and-the-Future-of-SEO.aspx>.
  26. Griffiths, R. T. (2006). History of the Internet. Najdeno 3. februarja 2009 na spletnem naslovu <http://www.internethistory.leidenuniv.nl/index.php3?c=7>.
  27. Interview with Matt Cutts on Search and SEO in China. Najdeno 15. aprila 2009 na spletnem naslovu <http://www.chinamyhosting.com/seoblog/2007/04/10/interview-matt-cutts-en/>.
  28. Iskold, A. (2006). Hakia – First Meaning – based Search Engine Najdeno 22. marca 2009 na spletnem naslovu [http://www.readwriteweb.com/archives/hakia\\_meaning-based\\_search.php](http://www.readwriteweb.com/archives/hakia_meaning-based_search.php).

29. Iskold, A. (2008, 29. maj). Read Write Web: Semantic Search: The Myth and Reality. Najdeno 29. januarja 2009 na spletnem naslovu [http://www.readwriteweb.com/archives/semantic\\_search\\_the\\_myth\\_and\\_reality.php](http://www.readwriteweb.com/archives/semantic_search_the_myth_and_reality.php).
30. Kartoo: About Kartoo. Najdeno 15. marca 2009 na spletnem naslovu <http://www.kartoo.net/eng/index.html>.
31. Knight, C. (2009). ReadWriteWeb. Najdeno 5. aprila 2009 na spletnem naslovu [http://www.readwriteweb.com/archives/top\\_100\\_alternative\\_search\\_engines\\_mar07.php?page=2](http://www.readwriteweb.com/archives/top_100_alternative_search_engines_mar07.php?page=2).
32. Lardinois, F. (2009). See Wolfram Alpha in Action: Our Screenshots. Najdeno 30. aprila 2009 na spletnem naslovu [http://www.readwriteweb.com/archives/see\\_wolfram\\_alpha\\_in\\_action\\_-\\_video\\_and\\_screenshots.php](http://www.readwriteweb.com/archives/see_wolfram_alpha_in_action_-_video_and_screenshots.php).
33. Liedtke, M. (2007, 30. maj). People Power Fuels New Search Engine. Najdeno 27. aprila 2009 na spletnem naslovu <http://www.sfgate.com/cgi-bin/article.cgi?f=/n/a/2007/05/30/financial/fl50126D41.DTL&type=business>.
34. MacManus, R. (2008). ReadWriteWeb: Interview with Google's Matt Cutts about Next-Generation Search. Najdeno 12. aprila 2009 na spletnem naslovu [http://www.readwriteweb.com/archives/interview\\_with\\_matt\\_cutts\\_next\\_generation\\_search.php](http://www.readwriteweb.com/archives/interview_with_matt_cutts_next_generation_search.php).
35. MacManus, R. (2009) ReadWriteWeb: What is a Search Engine - Now and Future. Najdeno 27. januarja 2009 na spletnem naslovu [http://www.readwriteweb.com/archives/what\\_is\\_a\\_search\\_engine.php](http://www.readwriteweb.com/archives/what_is_a_search_engine.php).
36. MacManus, R. (2009). ReadWriteWeb: AI Favored Search 2.0 Solution. Najdeno 14. januarja 2009 na spletnem naslovu [http://www.readwriteweb.com/archives/ai\\_favored\\_search20\\_solution.php](http://www.readwriteweb.com/archives/ai_favored_search20_solution.php).
37. Markoff, J. (2006, 12 november). Entrepreneurs See a Web guided by Common Sense Najdeno 3. aprila 2009 na spletnem naslovu <http://www.nytimes.com/2006/11/12/business/12web.html>.
38. McNichol, T. (2009). The Wales Rules for Web 2.0. Najdeno 5. junija 2009 na spletnem naslovu [http://money.cnn.com/galleries/2007/biz2/0702/gallery.wikia\\_rules.biz2/index.html](http://money.cnn.com/galleries/2007/biz2/0702/gallery.wikia_rules.biz2/index.html).
39. MG Rush. (2009). Glossary. Najdeno 11. aprila 2009 na spletnem naslovu <http://www.mgrush.com/content/view/70/33>.
40. Najdi.si. (2008). Izdelki in storitve. Najdeno 15. aprila 2009 na spletnem naslovu <http://www.najdi.si/about/products.html>.
41. Oblak – Črnič, T. (2008). O začetkih interneta na slovenskem. Najdeno 26. aprila 2009 na spletnem naslovu <http://www.javnost-thepublic.org/media/datoteke/oblak-crnic.pdf>.
42. O'Reilly, T. (2005). What is Web 2.0. Najdeno 13. aprila 2009 na spletnem naslovu <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>.

43. Perez, S. (2009). ReadWriteWeb: Google: We're Not Doing a Good Job with Structured Data. Najdeno 19. maja 2009 na spletnem naslovu [http://www.readwriteweb.com/archives/google\\_were\\_not\\_doing\\_a\\_good\\_job\\_with\\_structured\\_data.php](http://www.readwriteweb.com/archives/google_were_not_doing_a_good_job_with_structured_data.php).
44. Radhakrishnan, A. (2009). Search Engine Journal: 9 Semantic Search Engines That Will Change the World of Search. Najdeno 3. aprila 2009 na spletnem naslovu <http://www.searchenginejournal.com/semantic-search-engines/9832/>.
45. Roush, W. (2007). Technology Review: People-Powered Search For the legions of Internet users contributing to new "human-assisted search" sites, no job is too small. Najdeno 6. marca 2009 na spletnem naslovu [http://www.technologyreview.com/read\\_article.aspx?ch=specialsections&sc=search&id=18655](http://www.technologyreview.com/read_article.aspx?ch=specialsections&sc=search&id=18655).
46. Sherman, C. (2006). Search Engine Watch: What's the Big Deal With Social Search? Najdeno 22. aprila 2009 na spletnem naslovu <http://searchenginewatch.com/3623153>.
47. Sherrets, D. & Venture Beat (2008). Google's Marissa Mayer: Social search is the future. Najdeno 3. februarja 2009 na spletnem naslovu <http://venturebeat.com/2008/01/31/googles-marissa-mayer-social-search-is-the-future/>.
48. Sherrets, D. (2007). Facebook search expands — will it take on Google? Najdeno 5. junija 2009 na spletnem naslovu <http://venturebeat.com/2007/11/11/facebook-search-expands-will-it-take-on-google/>.
49. Spivack, N. (2007). The Third-Generation Web – Web 3.0 (2007). Najdeno 5. aprila 2009 na spletnem naslovu [http://www.intentblog.com/archives/2007/02/nova\\_spivack\\_th.html](http://www.intentblog.com/archives/2007/02/nova_spivack_th.html).
50. Sproose: About. Najdeno 5. junija 2009 na spletnem naslovu <http://sproose.com/static/en/corporate/about>.
51. Sullivan, D. (2000). Search engine Watch. Najdeno 22. aprila 2009 na spletnem naslovu <http://searchenginewatch.com/3505621>.
52. Sullivan, D. (2002). Search Engine Watch: Specialty Search Engines. Najdeno 22. aprila 2009 na spletnem naslovu <http://searchenginewatch.com/2156351>.
53. Sullivan, D. (2007, 2. februar). Google Ramps Up Personalized Search. Najdeno 15. maja 2009 na spletnem naslovu <http://searchengineland.com/google-ramps-up-personalized-search-10430>.
54. Sutter, J. D. (2009). New search engines aspire to supplement Google. Najdeno 5. junija 2009 na spletnem naslovu <http://edition.cnn.com/2009/TECH/05/12/future.search.engine/>.
55. Webb, M. (2001, 18. junij). Going With Open Source Software. Najdeno 4. aprila 2009 na spletnem naslovu <http://www.techsoup.org/learningcenter/software/archives/page9905.cfm>.



56. Wikitionary: Crowdsourcing. Najdeno 17. marca 2009 na spletnem naslovu <http://en.wiktionary.org/wiki/crowdsourcing>.
57. Wired: Net Surf. Najdeno 5. marca 2009 na spletnem naslovu [http://www.wired.com/wired/archive/1.05/net\\_surf.html](http://www.wired.com/wired/archive/1.05/net_surf.html).
58. Wood, P. (2009). Semi-Structured Data. Najdeno 25. aprila 2009 na spletnem naslovu <http://www.dcs.bbk.ac.uk/~ptw/teaching/ssd/slide3.html>.



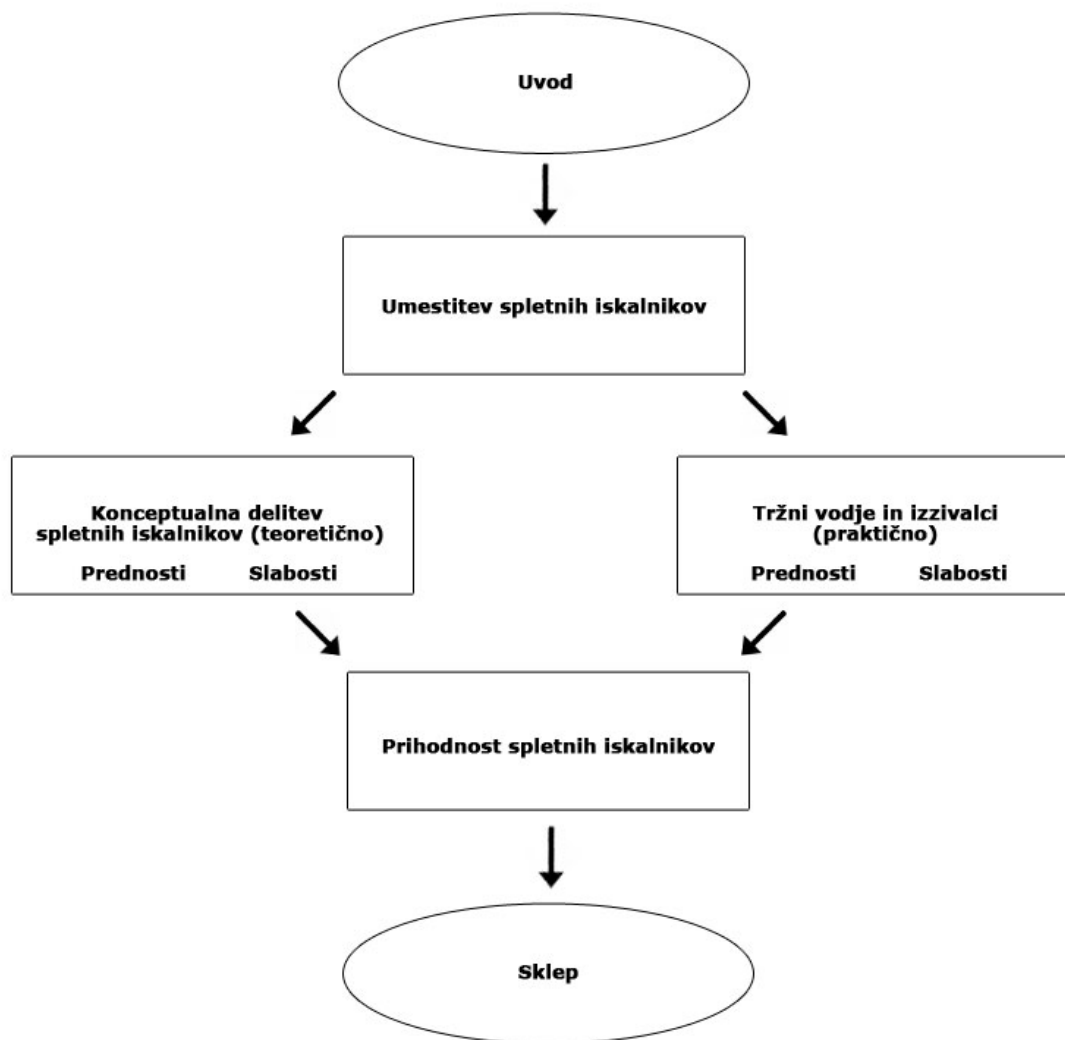
## Priloga 1

### ANGLEŠKO-SLOVENSKI SLOVARČEK UPORABLJENIH IZRAZOV

Information Highway	informacijska avtocesta
Net	mreža
Matrix	matrica
Web Site	spletišče, spletno mesto
Web page	spletna stran, dinamična spletna stran
Web	splet
World Wide Web ali WWW	svetovni splet
Hyperlink	hipertekstovne povezave
URL (uniform resource locator)	internetni naslov, na katerem se nahaja vsebina
Relevance feedback	povratna zveza relevantnosti
GPL – GNU public licence	javna licenca
Indexing	indeksiranje
Web crawling	spletno plazenje
Crawler, spider	Pajek (črv) je program, ki samodejno poišče povezave na druge spletne strani in jim sledi, zaradi česar lahko poišče veliko število spletnih strani.
Peer to Peer ali P2P	vsak z vsakim
Social networking	socialno druženje

Web 2.0	splet druge generacije
HTTP (Hypertext Transfer Protocol)	protokol za izmenjavo besedil ter grafičnih, zvočnih in drugih večpredstavnostnih vsebin na spletu
Crowdsourcing	uporaba množice ljudi
Visual search engine	vizualni spletni iskalnik
Tag cloud	oblak, množica značk
Clustering	gručenje
Typo	slovnična nepravilnost
Deep web	globok splet
Bookmarks	zaznamki
Tag	označevalec ali značka
Voting	glasovanje
Computational knowledge engine	računalniški iskalnik znanja


Slika 4: Koncept diplomske naloge



Vir: Lasten vir, 2009.

Priloga 3

Slika 5: Primer iskanja s spletnim iskalnikom Wolfram Alpha

 **WolframAlpha**™ computational... knowledge engine

=


**Input interpretation:**  
Munich, Bavaria, Germany to New York, United States

**Distance:** [Show non-metric units](#)

	distance	flight time
Munich to New York	6500 km (kilometers)	7 hours 20 minutes

(assuming direct flight path at 890 km/h)

**Path:**



**Current local times:**

Munich, Bavaria	10:00 am CEST   Tuesday, May 19, 2009
New York	4:00 am EDT   Tuesday, May 19, 2009

**Populations:**

	city population	metro area population
Munich, Bavaria	1.26 million people	
New York	8.143 million people	18.61 million people

Vir: WolframAlpha, 2009.