

Univerza v Ljubljani

Filozofska fakulteta

Oddelek za bibliotekarstvo, informacijsko znanost in založništvo

*Spletni iskalniki:
prikazovanje rezultatov iskanja*

Mentor: doc. dr. Jure Dimec

Avtorji: Nina Klemenčič
Marija Lavrih
Mateja Magdič
Mihela Pauman
Maja Vratuša

Ljubljana, maj 2009

Kazalo

UVOD	- 3 -
1. KAKO DELUJEJO SPLETNI ISKALNIKI?.....	- 3 -
2. KAKO ISKALNIKI RAZVRŠČAJO ZADETKE?	- 3 -
3. IZBRANI ISKALNIKI	- 4 -
• GOOGLE	- 4 -
• YAHOO	- 6 -
• MICROSOFT LIVE SEARCH	- 7 -
• CUIL.....	- 8 -
• POWERSET	- 9 -
4. PRIMERJAVA ISKALNIKOV	- 10 -
ZAKLJUČEK.....	- 11 -

UVOD

Internet je za večino ljudi postal prvi vir, na katerega se obrnejo, kadar iščejo informacije o določenem izdelku ali storitvi. Spletni iskalniki so orodje, brez katerega danes praktično ne gre. So tista spletna stran, kjer se zelo pogosto začne in odvija naše bivanje v internetu. Uporabljamo jih verjetno vsak dan, toda že skoraj instinktivno, brez pretiranega razmišljanja, kaj vse znajo ter kako jih uporabljati. Toda zanimivo, prav orodje, ki ga uporabljamo vsak dan, je za mnoge v resnici precej neznano. Večina uporabnikov namreč iskalnike uporablja le v njihovi osnovni obliki, saj so zadetki navadno povsem zadostni. Ko pa kljub naporom z različnimi ključnimi besedami ne najdejo iskanega, večina odneha, čeprav bi lahko preverili možnosti naprednega iskanja, dodatne storitve iskalnika ali pa posegli po drugem, bolj specializiranem iskalniku.

1. Kako delujejo spletni iskalniki?

Iskalniki za pregledovanje vsebine uporabljajo posebne programe - spletne robote ali pajke. Spletni pajki pregledujejo spletne strani tako, da sledijo vsem povezavam (linkom) na posamezni podstrani.

Ko je cikel obiskovanja vseh strani v spletu končan, se s pomočjo posebnih programov naredi indeks dokumentov v iskalniku (indeksiranje). Iskalnik rezultate iskanja za vsako iskalno frazo razvrsti po posebej določenem računalniškem algoritmu. Vsak iskalnik ima svoj algoritem razvrščanja zadetkov.

Rezultati iskanj se prikazujejo med zadetki, vse dokler ni končan naslednji cikel osveževanja indeksa. Takrat se rezultati obnovijo. V povprečju se to zgodi enkrat tedensko ali 3-4 krat mesečno. Zato je potrebno vedeti, da se vsaka sprememba, ki jo naredimo na spletni strani, ne zazna takoj v iskalniku, temveč šele ob naslednjem osveževanju indeksa.

2. Kako iskalniki razvrščajo zadetke?

- Indeksiranje:

Ko spletni pajki (roboti) končajo s ciklom obiskovanja spletnih strani, se s pomočjo posebnih algoritmov naredi indeks dokumentov v iskalniku.

Iskalnik rezultate iskanja za vsako iskalno frazo razvrsti po posebej določenem računalniškem algoritmu. Vsak iskalnik ima svoj algoritem oz. pravila, po katerem razvršča zadetke.

- Ali so pravila rangiranja znana?

Kriteriji rangiranja pri posameznih iskalnikih so skrivnost, ki jo iskalniki zaradi možnosti zlorabe skrivajo. Iskalniki zadetke razvrščajo s pomočjo algoritma, nikoli ni višje uvrščena

stran, ki je za to plačala (razen spozoriranih povezav, ki so na vrhu zadetkov in so tudi posebej označene).

- Več kot 200 pravil za rangiranje:

Iskalniki za razvrščanje svojih zadetkov upoštevajo preko 200 različnih faktorjev. Pravila niso znana, kako torej vemo, da jih je toliko in katera so...

Optimizacija spletnih strani je dejavnost, kjer do večine podatkov in pravil pridemo na podlagi iskustvenih metod. Na podlagi lastnih in tudi na osnovi preiskovanj drugih, ki se s to dejavnostjo ukvarjajo. Številni forumi na domačih in tujih internetnih straneh so vir koristnih informacij.

- Iskalniki pravila redno spreminjajo:

Iskalniki svoja pravila redno spreminjajo. Težo dajo drugim faktorjem, veliko pravil zgubi na svoji veljavi. Veliko faktorjev, ki so še pred dvema letoma imeli precejšen vpliv na uvrstitve v iskalnikih, danes ne igrajo več bistvene vloge pri razvrščanju zadetkov.

3. IZBRANI ISKALNIKI:

- **Google** (<http://www.google.si>)



- ***Triki, nasveti, operatorji za spletni iskalnik Google***

Tukaj so razložene različne vrste iskanj z najbolj popularnimi operatorji, ki so na voljo za uporabo pri iskalniku Google.com.

Prva so razložena splošna iskanja. Potem lahko Google.com uporabljamo tudi kot preprost kalkulator. Na koncu so pa še napredni operatorji.

▪ **Splošna iskanja na google.com**

PRIMER ISKANJA	ZADETKI
lep dan	vsebujejo besedo lep in dan
lep OR dan	vsebujejo besedo lep ali dan
“ali je lep dan“	vsebujejo točno tako frazo
lep -dan	ne vsebujejo besedice dan
Vojna Zvez +5	nujno vsebujejo tudi številko 5
~lep dan	iskanje po podobnih besedah (sinonimih).
define: computer	definicija besedice computer.
lep * dan	med besedama se nahaja ena ali več besed

▪ **Google.com kot kalkulator**

OPERATOR	POMEN	PRIMER UPORABE
+	Prištevanje	45 + 39
-	Odštevanje	45 ? 39
*	Množenje	45 * 39
/	Deljenje	45 / 39
% of	odstotek od	45% of 39
^	Potenciranje	2^5

▪ **Napredni operatorji na Google.com**

OPERATOR	POMEN	PRIMER UPORABE
site:	Išči samo po enem spletnem mestu	google site:ruph.sopca.com (Išči google po spletnem mestu ruph.sopca.com)
[#]...[#]	Išči v intervalu števil	DVD player \$100..150 (Išči za DVD playerje med \$100 in \$150)
date:	Išči samo v določenih mesecih	volitve date: 3 (Išči volitve v zadnjih 3 mesecih; google omogoča še uporabo 6 in 12 mesečne restrikcije)
link:	Referenčne povezave	link:www.google.com (Google pokaže del spletnih strani, ki imajo povezavo na www.google.com.)
info:	Informacije o spletnem mestu	info:www.google.com (Informacije o spletnem mestu www.google.com.)
related:	Sorodne strani	related:www.google.com (Prikaže podobne strani, kot je www.google.com.)

- **Yahoo** (<http://www.yahoo.com>)



Na vrhu strani www.yahoo.com imamo iskalno polje, kamor vpišemo iskani pojem. Spletni iskalnik nam pokaže 10 zadetkov na eno stran. Zadetki so uporabniku predstavljeni na razumljiv način, saj so naslovi določene spletne strani poudarjeno napisani z modro barvo, pod tem se skriva krajši opis spletne strani (v črni barvi) in nato sledi še URL naslov, ki je zapisan z zeleno barvo. Predvsem z barvami vplivajo na večjo preglednost.

Uporabnik spletnega iskalnika Yahoo pa lahko svoje iskanje omeji po določenih kategorijah, saj ima možnost izbire, da se zadetki pokažejo kot slike, videi, itd. Tako si uporabnik sam izbere katere vrste zadetkov naj bodo prikazani. Yahoo vse svoje zadetke razvršča po določenemu algoritmu, le pri »News search«, si lahko uporabnik izbere ali naj bodo zadetki razvrščeni po njihovem algoritmu ali po datumu. Ker pa vsak posameznik drugače zazna razvrščenost zadetkov, sem naredila 3 primere, kjer sem poskušala ugotoviti ali so na vrhu res predstavljeni najbolj relevantni zadetki za večino uporabnikov.

- **Primer 1:**

Pri prvem primeru me je zanimala vremenska napoved v Sloveniji, zato sem v iskalno polje vpisala »vreme po Sloveniji« in dobila sem 1,870,000 zadetkov 0.02 sekundah. Ker je nemogoče pregledati vse zadetke, sem pregledala pri dve strani (torej 20 zadetkov). Za mene je bilo relevantnih le 7 zadetkov, ostali pa so bili čisto neprimerni, saj iz njih nebi mogla izvedeti želenega podatka.

- **Primer 2:**

V drugem primeru sem dobila podatek o rodnosti v Sloveniji, zato sem zopet v iskalno polje vpisala »rodnost v Sloveniji« in dobila sem 15,300 zadetkov v 0.38 sekundah. Ponovno sem pregledala le dve strani zadetkov. Za mojo informacijsko potrebo je bilo 5 zadetkov od 20 relevantnih, ostali podatki mi niso koristili.

- **Primer 3:**

V zadnjem primeru sem poskušala dobiti kje so kolesarske poti po Sloveniji, in sem v iskalno polje vpisala »kolesarske poti po Sloveniji« ter dobila 68,700 zadetkov v 0.28 sekundah. Želela sem izvedeti kje vse so kolesarske poti in sem zopet pregledala le prvi dve strani zadetkov. Po

pregledu zadetkov sem dobila le 4 dokaj ustrezne spletne strani, a vseeno nisem prišla do želene informacije.

V vseh treh primerih sem ugotovila, da vsi relevantni zadetki niso bili na vrhu, zato je potrebno pregledati več strani zadetkov in poskušati tudi z različnimi ključnimi besedami. Vsekakor pridejo v pomoč tudi operatorji, s katerimi uporabnik sam omeji rezultate na nekoliko bolj relevantne ali pa s tem zmanjša število zadetkov.

- **Microsoft Live Search (<http://www.live.com>)**



- **Primer 1:**

V prvem poskusu sem želela poiskati nekaj zgodovinskih dejstev o gradu Grad na Goričkem. Dobila sem sicer okrog 1200 zadetkov, vendar na nekaj prvih straneh nisem našla ustreznih informacij. Večinoma so bili zadetki povezave do zemljevidov oz. so želeli pokazati, kje se ta grad nahaja.

- **Primer 2:**

Drugo poizvedbo sem naredila na temo jagodna marmelada. Dobila sem 101 zadetek. Na prvi strani sem dobila le en relevanten zadetek. Na neuspešnost je vplivala tudi pomanjkljiv iskalni niz. Ker so me zanimali recepti za marmelado, bi morala dodati še besedo »recepti«.

- **Primer 3:**

Ker nam iskalnik Live Search omogoča tudi iskanje slik, sem v zadnji poizvedbi iskala slike. V iskalno okence sem vpisala, da me zanimajo slike o zlatih prinašalcih. Iskalnik mi je ponudil 116 zadetkov. Že ob hitrem preletu sem ugotovila, da se med slikami zlatih prinašalcev pojavljajo še slike drugih pasem in razni napisi.

Iskalnik nam na eni strani ponudi deset zadetkov. Med posameznimi zadetki je dovolj prostega prostora, da je stran pregledna. Posamezni zadetek pa za lažjo preglednost vsebuje več barv. Naslov spletne strani je modre barve in posamezne besede (iskalni niz besed) so še posebej poudarjene. Tudi velikost pisave je večja, da naslov zlahka razločimo od ostalih

podatkov. Naslovu spletne strani sledi opis spletne strani, ta je zapisan v črni barvi in manjši pisavi, kakor naslov. Pod opisom strani najdemo URL naslov, ki pa je obarvan zeleno in zato še posebej izstopa. Barve in razporeditev zadetkov nam zagotavljajo zadostno preglednost. Zanimivost so tudi zeleni znaki ob naslovu spletne strani, ki potrjujejo, da je spletna stran varna.

- **Cuil** (<http://www.cuil.com>)



Pri odprtju strani je prikazano polje kamor vpišemo besedo. Iskalnik je angleški, tako da moramo besede pisati v angleškem jeziku, vendar nam najde besede tudi če jih napišemo v slovenskem jeziku. Prav tako na tem iskalniku najdemo povezave do slovenskih spletnih strani. Odvisno je v katerem jeziku vpišemo iskalni pojem. Iščemo oz poizvedujemo lahko na 124,426,951,803 spetnih straneh. Pri iskanju se nam pojavijo pomožno oz. podokno, ki nam ponudi izbiro med naprednim iskanjem in iskanjem po kategorijah. Pri vsakem načinu se pojavi več primerov, ki vsebujejo besede, ki smo jih vpisali.

- **Primer 1:**

Pri prvem iskanju sem vtipkala besedo library (knjižnica). Pred začetkom iskanja se pojavi okence kjer mi ponuja izbro med možnostmi: library house (knjižnična hiša), onlyn library (internetna knjižnica), library science (knjižničarska znanost), library jurnal (knjižnični časopis), kategorije; bibliotekrstvo in informacijska znanost, bibliotekrsko klasificiranje in katalogiziranje, ALA (American library asociation)... Nisem izbrala ničesar izmed njih ampak sem iskalna na splošno knjižnice. Dobila sem 3,084,786,172 rezultatov. Pregledala sem prvih 7. Prvi dva rezultata sta bila knjižnici dveh univerz (York) nato pa so sledile povezave do splošnih knjižnic. Pri vsakem naslovu je bil kratek opis njihove iskalne baze ter povezava knjižnice do drugih podatkov. Ko sem kliknila na katerega izmed zapisov oz. zadetkov sm dobila povezavo do iskalnikov oz. baz podatkov knjižnice.

- **Primer 2:**

Izbrala sm Hollywood. Dobila sem 1,559,459,629. Prvih nekaj jih opisuje kaj je Hollywood, filmsko industrijo, mesto Hollywood (leži na Floridi)... Tudi pri tem poizvedovanju imajo

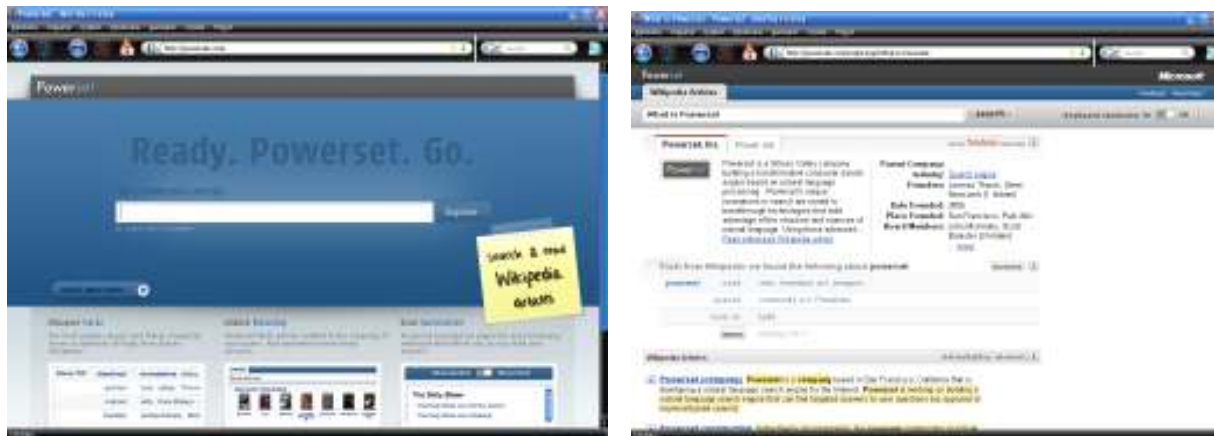
rezltati oz. zadetki kratek komentar oz. opis, ki nam na kratko opiše o čem piše komentar, oz. kaj bomo lahko videli in prebrali če kliknemo na to povezavo.

- *Primer 3:*

Izbrala sem pojem kremšnite. Vpisala sem slovesko besedo saj sem želela izvedeti ali bo sploh prikazalo kaj zadetkov. Dobila sem 11,662 zadetkov. Po večini so bile povezave navezane na različne slovenske spletne strani od mladine, do različnih blogov, kjer so ponujali recepte, ali pa jih komentirali...

Pri vseh primerih sem opazila, da je pri vsake zadetku komentar (opis), pod njim pa spletna povezava, ki ti ponudi še več informacij in vpogledov do stvari o kateri smo izvedli poizvedovanje.

- ***Powerset*** (<http://powerset.com>)



Na vrhu strani imamo iskalno polje, tja vpišemo iskani pojem. Iskalnik nam poda 10 zadetkov na eno stran. Med zadetki je najprej definicija iskanega pojma, potem pa ostali zadetki, ki so predstavljeni razumljivo. Sam iskani pojem je obarvan modro, ko pa se pojavi v opisu je obarvan rumeno. Njegova pomanjkljivost pa je da nam ne izpiše, koliko je vseh zadetkov skupaj.

- *Primer 1:*

Zanimal me je projekt svetovne digitalne knjižnice, zato sem v iskalno polje vtipkala Digital library (digitalna knjižnica). Med zadetki so poleg iskanega pojma podane digitalne knjižnice nekaterih držav. Na koncu strani so nam podane še povezave do sorodnih strani v zvezi z iskanim pojmom. Zadetkov je sicer veliko, vendar iskani pojem opisujejo zgolj prvi trije, ostali pa zadevajo državne digitalne knjižnice.

- *Primer 2:*

V tem primeru me je zanimalo taljenje ledu na Antarktiki, vtipkala sem Melting of Antarctica. Ker je vseh zadetkov ogromno, sem pregledala zgolj prve 3 strani, vendar je bilo zame relevantnih samo prvih 5 zadetkov, ostali niso bili primerni. Ravno tako so tudi tukaj na koncu strani dodane povezave sorodnih pojmov.

- *Primer 3:*

V tem primeru me je zanimalo, kako hitra prehrana vpliva na otrokovo vedenje. Zadetki opisujejo predvsem psihologijo otroškega vedenja in definicije hitre prehrane, tako da niso relevantni. Dodane so povezave na članke sorodne tematike.

4. PRIMERJAVA ISKALNIKOV:

Statistični podatki kažejo, da področje iskanja obvladujejo predvsem trije največji sistemi – Google, Yahoo in Microsoft (Live Search). Po podatkih družbe Nielsen NetRatings ima Google globalno 53,6 % vseh iskanj v spletu, Yahoo 19,9 % in Microsoft 12,9 %.

Google je konkurenco hitro prekosil predvsem zaradi inovativnega algoritma PageRank, ki zagotavlja razvrstitev zadetkov po številu povezav nanje, torej po priljubljenosti (v resnici pa Google celotne zgodbe okoli strategije iskanja nikoli ne razkriva in jo nenehno dopolnjuje). Google je eden tistih, ki imajo najbolj razdelano napredno iskanje, kjer lahko izbiramo kriterije in omejujemo domeno iskalnih izrazov. Iščemo lahko po jezikih, regiji, domenah, tipih datotek, datumu zadnje spremembe strani in seveda poljubnih ključnih besedah. Ena izmed manj znanih možnosti Googla je iskanje po tujih straneh s prevodom zadetkov v slovenščino. V zadnjem času je še posebej zanimiva povezav s sistemom za samodejno prevajanje Google Translate, kjer lahko iščemo besede v enem jeziku, rezultate pa nato prikažemo v drugem (prevedenem) jeziku. Tega tekmeči večinoma ne ponujajo, zagotovo pa ne z možnostjo prevodov v slovenščino.

Bili so časi, ko je bil Yahoo najboljši spletni iskalnik, Google pa zgolj eksperiment dveh študentov, ki sta si zamislila nekaj novega. Po vseh teh letih so se stvari spremenile, Yahoo pa se je iz spletnega iskalnika razvil v veliko korporacijo, ki razvija in ponuja številne druge spletne storitve, zaradi katerih je sam iskalnik skoraj nekoliko preveč zbledel. Yahoo je v zadnjih letih imel nekoliko konfuzno strategijo, zaradi katere so zabredli v težave in postali tarča morebitnega prevzema. Še svež je spomin na prizadevanja Microsofta, da bi jih prevzel. Sam iskalnik se predstavi z obilico informacij, ravno nasprotno kot Google, ki ima na tem mestu povsem minimalističen uporabniški vmesnik. Rezultati iskanja so v osnovi predstavljeni podobno kot drugod, imamo pa več možnosti za preiskovanje po različnih virih informacij, skupaj z video posnetki, fotografijami in zvoki. Naročiti se je mogoče tudi na rezultate točno določenih ponudnikov informacij (npr. novic).

Microsoft si že lep čas močno prizadeva, da bi si s svojim spletnim iskalnikom odrezal večji tržni delež, zlasti zaradi dobičkonosnega oglaševanja. Tudi zaradi tega so pred nekaj leti dotedanji iskalnik v okviru spletnih povezav MSN osamosvojili in poimenovali Live Search. Iskalnik ima vse, kar imajo najbližji tekmeči, čeprav sami rezultati iskanja včasih niso dovolj relevantni. Live Search se hitro razvija in po količini različnih specializiranih iskalnikov morda

celo prekaša večje tekmece. Poleg spletnih strani lahko preiskujemo fotografije, video posnetke, novice, zemljevide, lokalne vsebine, izdelke, zdravstvene informacije, cene letalskih vozovnic, iščemo lahko celo po priljubljenih vsebinah drugih iskalcev. Posebnost, ki je ne najdemo povsod, je podajanje odgovora na neposredna vprašanja. Povedati zna, recimo, katero je glavno mesto posamezne države, podati definicijo kratice, celo izračunati rezultat enačbe in opraviti pretvorbo med merskimi enotami.

Prevladujoči omenjeni trije iskalniki pa prav gotovo niso vsi. V spletu nenehno nastajajo novi projekti, ki se pogosto specializirajo za posamezno področje ali pa skušajo rezultate prikazati drugače kot vodilni ponudniki.

Lani predstavljen iskalnik Cuil sta ustanovila dva nekdanja zaposlena iz Googla. Pozornost je vzbudil predvsem s trditvijo, da so največji spletni iskalnik na svetu, ki ima indeksiranih čez 120 milijard strani po vsem svetu. To je več kot pri Googlu. Podobno kot drugi ima minimalistično prvo stran, pri prikazu rezultatov pa je bolj radodaren. Zadetki so opremljeni z daljšimi opisi, pogosto so stranem dodane manjše slike, celota je oblikovana kot novičarske povezave. Najbolj verjetne alternative ali variacije iskanega izraza združi v lasten zavihek, in to nam lahko skrajša iskanje. Žal rezultati iskalnika Cuil pogosto niso tako relevantni kot v primeru Googla. Iskalnik je pač nov in se še razvija. Cuil med vsemi iskalniki prikaže najlepše oblikovane zadetke.

Vse več je iskalnikov, ki namesto seznama spletnih zadetkov ponujajo tudi odgovore na konkretna vprašanja. Najbolj znan med njimi ta hip je nedvomno Powerset, ki na vrhu zadetkov prikaže konkretne odgovore. Ti so zasnovani pretežno na člankih Wikipedie, vendar ne samo to. Zanimiva je tudi možnost pregledovanja zadetkov v mini pregledovalniku, kjer so izpostavljene ključne in alternativne besede. Powerset zna odgovoriti na številna vprašanja, ne samo iskati po številnih ključnih besedah. Mnogi menijo, da lahko doseže velik uspeh in se čez čas celo kosa z Googlom.

ZAKLJUČEK

Spletni iskalniki najdene informacije navadno predstavijo kot sezname rezultatov, ki jim pravimo zadetki. Sam način predstavitve zadetkov je ena tistih stvari, kar posamezen iskalnik loči od drugih. Zmotno je tudi misliti, da spletni iskalniki danes iščejo samo po spletnih straneh. Danes je to le eden, čeprav poglaviti vir. Iščemo lahko tudi po zbirkah podatkov, imenikih in po drugih internetnih entitetah (npr. novičarske skupine).

Slaba stran iskalnikov je seveda to, da nam praviloma dajo preveč zadetkov. Poplava informacij, po možnosti brez kake posebne strukture, je seveda približno tako koristna, kot če iskalnika nimamo. Zato pridejo na pomoč razne tehnike in algoritmi za razvrščanje podatkov po različnih kriterijih. Najbolj iskan in ponavadi privzet način je razvrstitev po relevantnosti. Prav razvrščanje po relevantnosti je področje, na katerem se iskalniki zelo razlikujejo, saj vsak stavi na svoj princip.

Iskalniki so nekaj, kar se nenehno spreminja. Pa ne le njihov indeks vselej ažurnih vsebin na milijardah spletnih strani po vsem svetu. Spreminja se tudi njihova zasnova, tako ogrodje kot podoba. Avtorji nenehno pilijo algoritme, s katerimi bi dali še bolj relevantne rezultate z minimalno količino ključnih besed oziroma truda.