



# Vještine komuniciranja u inženjerstvu

## Pretraživanje informacija



**Juraj Petrović**



# Google

## UPOZNAJMO GOOGLE

Tražilica	Tržišni udio (siječanj 2022)
Google	92,01%
Bing	2,96%
Yahoo!	1,5%
Baidu	1,17%

# Što je Google?

- osnovali 1998. Larry **Page** i Sergey **Brin**
  - doktorski **studenti** na **Stanford sveučilištu**



- misija:

*“organizirati ljudsko znanje  
i omogućiti univerzalni pristup i korisnost”*

- vrijednost oko **1,8 bilijuna USD**
- **+180000** zaposlenika (Alphabet Inc.)
- tražilica s cijelim spektrom različitih **aplikacija i alata**



# Google



Google Search

I'm Feeling Lucky

Google offered in: [hrvatski](#)

# Kako Google radi?



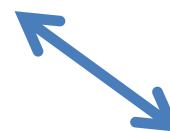
## Web pauci (spider, crawler)

Tražilica pokreće  
**web pauke** da nađu  
mrežno mjesto  
koje treba indeksirati



## Program za indeksiranje

prikuplja, organizira i  
pohranjuje materijale  
kako bi **pretraživanje**  
bilo brzo i lagano



**Program za  
pretraživanje**  
omogućava korisniku  
**unošenje upita** i  
isporučuje rezultate  
temeljene na indeksu



# Osnovno pretraživanje - Google

- **svaka riječ** je važna
  - redoslijed riječi je važan
  - najviše 32 riječi
  - napišite što vas zanima koristeći **što manje riječi**
- pitajte se: “Kako računalo razmišlja?”
  - iako Google radi sve da nauči kako vi razmišljate
- izraz **u navodnicima** se **doslovno** traži
  - "hotel Zagreb" ≠ hotel Zagreb
  - automobil ≠ "automobil"



- **OR logički operator:**  
maraton OR utrka
- **isključivanje** izraza (minus):  
virus -software
- **nadomještanje** čitavih riječi (\*):  
"najstarija \* na svijetu"
- **iscrpni popis:**  
<https://support.google.com/websearch/answer/2466433?hl=en>

# Posebni filtri (filetype, site)

- filetype:
  - koristite **pretraživanje** određene vrste **datoteka**
    - **pdf** ili **doc** za vladina i druga izvješća, službene dokumente
    - **xls** za podatke i statistike
    - **ppt** ili **pdf** za prezentacije

robotika (filetype:ppt OR filetype:pptx)

- site:
  - **suzite pretraživanje na domenu**
    - „strojno učenje” site:fer.hr
    - linux tečajevi site:srce.hr



# Posebni filtri (before/after)

- before/after:
  - **vremensko ograničavanje** rezultata pretrage

web development after:2020

covid before:2019-10-25

# Google napredno pretraživanje

## Advanced Search

### Find pages with...

all these words:

this exact word or phrase:

any of these words:

none of these words:

numbers ranging from:

to

### To do this in the search box.

Type the important words: tri-colour rat terrier

Put exact words in quotes: "rat terrier"

Type OR between all the words you want: miniature OR standard

Put a minus sign just before words that you don't want:  
-rodent, -"Jack Russell"

Put two full stops between the numbers and add a unit of measurement:  
10..35 kg, £300..£500, 2010..2011

### Then narrow your results by...

language:

any language

Find pages in the language that you select.

region:

any region

Find pages published in a particular region.

last update:

anytime

Find pages updated within the time that you specify.

site or domain:

Search one site (like wikipedia.org) or limit your results to a domain like .edu, .org or .gov

terms appearing:

anywhere in the page

Search for terms in the whole page, page title or web address, or links to the page you're looking for.

SafeSearch:

Show most relevant results

Tell [SafeSearch](#) whether to filter sexually explicit content.

file type:

any format

Find pages in the format that you prefer.

usage rights:

not filtered by licence

Find pages that you are free to use yourself.

Advanced Search

# Razne pomoći

- *search suggestions*
  - rezultati se dinamički pojavljuju tijekom unošenja upita
    - predviđa sadržaj upita i nudi različite mogućnosti

Google

I hate it when

i hate it when i'm studying and a velociraptor  
i hate it when jesus rides dinosaurs  
i hate it when  
i hate it when i'm studying and a velociraptor throws  
i hate it when he does that  
i hate it when i'm making a milkshake and  
i hate it when you  
i hate it when that happens  
i hate it when girls die  
i hate it when i lose my

Google Search

I'm Feeling Lucky

- zašto je bitan broj rezultata?

Google

"machine learning"

All

Images

Videos

News

About 165,000,000 results (0.59 seconds)

# Knowledge graph

- *knowledge graph* = sažeti prikaz informacija o rezultatu pretrage

Google

new york

All Images Maps News Videos More Settings Tools

About 6,100,000,000 results (1.50 seconds)





**New York City - Wikipedia**  
[https://en.wikipedia.org/wiki/New\\_York\\_City](https://en.wikipedia.org/wiki/New_York_City)



The City of **New York**, usually referred to as either **New York City** (NYC) or simply **New York** (NY), is the most populous city in the United States. With an ...

**Largest borough by population:** Brooklyn (2,63...)  
**Area code(s):** 212/646/332, 718/347/929, 917  
**Largest borough by area:** Queens – 109 square  
**Named for:** James, Duke of York

History of New York City · New York, New York · Geography of New York City · Pace

**Top things to do in New York**

 <b>Statue of Liberty</b> National Monument American icon in New York Harbor	 <b>Central Park</b> Urban oasis with ballfields & a zoo	 <b>Times Square</b> Bright lights & Broadway shows	 <b>Empire State Building</b> 103-story landmark with observatories
--	---	--	---

  
Map data ©2019 Google

## New York

City in New York

New York City comprises 5 boroughs sitting where the Hudson River meets the Atlantic Ocean. At its core is Manhattan, a densely populated borough that's among the world's major commercial, financial and cultural centers. Its iconic sites include skyscrapers such as the Empire State Building and sprawling Central Park. Broadway theater is staged in neon-lit Times Square.

**Local time:** Wednesday 05:15  
**Weather:** 1°C, Wind N at 6 km/h, 97% Humidity  
**ZIP Codes:** 100xx–104xx, 11004–05, 111xx–114xx, 116xx  
**Population:** 8.623 million (2017)  
**GDP (City, 2015):** US\$807 billion (1st)

New York travel guide

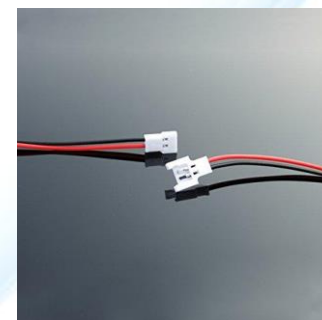
# Pronalaženje specifičnih podataka

- Vremenska prognoza  
**weather** zadar
- Lokalno vrijeme  
**time** new york
- Kalkulator  
 $\cos(\pi/4) * e^{\ln 2}$
- Riječnik  
**define** trepan
- Pretvaranje jedinica  
322 cm in inches
- Pretvaranje valuta  
100 usd in hrk  
100 btc in eth



# Google pretraga slika

- <https://images.google.com>
  - slike vezane za pretragu
  - pretraga po zadanoj slici
    - gdje se koristi **ova** slika?
    - gdje se koristi **moja** slika?
    - kako se zove to što je na slici (na stranom jeziku)?
    - je li slika izvorna ili preuzeta od negdje?



- algoritam je tajni
- Google rangira stranice na temelju:
  - svojstva stranice, primjerice:
    - ime domene, ključne riječi u domeni i poddomenama, ...
    - metapodaci stranice, gustoća ključnih riječi, vrijeme učitavanja stranice, korištenje stranice, ...
    - povratne veze (engl. *backlinks*)
    - ...

- daljnji faktori koji utječu na rangiranje:



- svojstva korisnika

- lokacija (IP adresa, GPS, ...)
- povijest pretraživanja
- uređaj i svojstva web preglednika (*browser fingerprinting*)

- Politički, ekonomski, društveni faktori

- dječja pornografija ili piratizirani sadržaji
- cenzura temeljena na državnim zakonima
- popularnost sadržaja

- Rezultat pretraživanja je ono na što će korisnik najvjerojatnije kliknuti
  - a NE najtočniji ili najažurniji izvor informacija
- „intelektualna izolacija”
  - rezultati pretraživanja i preporuke društvenih mreža i pretraživača su samo oni koji odgovaraju svjetonazorima korisnika



- brojne usluge
  - pretraživanje, YouTube, Android, Google Apps,...
- većina prihoda (oko 80%) dolazi od ciljanog oglašavanja koje omogućava profiliranje korisnika





**POSTOJI LI JOŠ NEŠTO OSIM GOOGLEA?**

# Osim Google-a ...

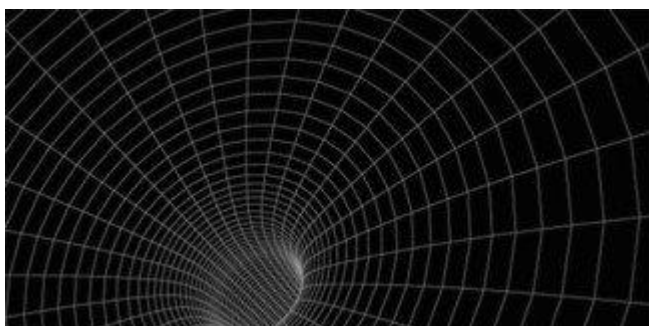
- druge tražilice:



Tražilica	Tržišni udio (kolovoz 2021)
Google	92,03%
Bing	2,48%
Yahoo!	1,5%
Baidu	1,39%

- meta-tražilice:

- **Dogpile** (Google, Yahoo!, Yandex, Bing)
- **Info.com** (Google, Yahoo!, Ask, Yandex, Open Directory, Kanoodle, LookSmart, ...)
- **Metacrawler** (Google, Yahoo!, Bing (formerly Live Search), Ask.com, About.com, MIVA, LookSmart)



# DEEP WEB VS. SURFACE WEB

# Što Google i pretraživači ne vide?

- **Ne vide:**

- sadržaje čiji **URL** nigdje **nije naveden**
- **baze podataka** i **‘dinamični sadržaj’**
- sadržaj **zaštićen zaporkom**
- sadržaj **zaštićen protokolom** koji sprječava indeksiranje
  - **/robots.txt**  
User-agent: \*  
Disallow: /
- pretplaćene sadržaje i **sadržaje koji se naplaćuju**



- *Surface web*
  - web sadržaj vidljiv pretraživačima
    - Wikipedia, news portali, društvene mreže, stranice institucija, web shopovi, ...
- *Deep web*
  - web sadržaj NEvidljiv pretraživačima
    - sadržaji zaštićeni lozinkama: baze podataka (medicinskih zapisa, znanstvenih radova, paywall sadržaji), sadržaji čiji URL nije nigdje objavljen (npr. javne poveznice na sadržaje u SRCE MojOblak)
- *Dark web*
  - dio *deep weba* dostupan u mrežama koje koriste Internet infrastrukturu, ali ju nadograđuju i trebaju posebne alate/konfiguraciju/autorizaciju za pristup, npr. Tor





- obrada prirodnog jezika
- predviđanje potreba korisnika iz senzora, analize prikupljenih podataka, ...
- digitalni osobni asistent



# A PRETRAGA STRUČNE ILI ZNANSTVENE LITERATURE?

- pronalaženje **publiciranih radova**
  - znanstveni
  - sve discipline
  - članci u časopisima, disertacije, knjige ...
- akademski izdavači, strukovne udruge, online repozitoriji, sveučilišta i dr.
- neki rezultati imat će na raspolaganju i **cjeloviti tekst**
- >100 milijuna tekstova na engleskom jeziku (~90%)
- alternativa: CiteSeer



- Pretraživanje po
  - autoru
  - naslovu
  - redanje po godini (**datumu**)
  - povezani, **slični radovi**
  - **publikaciji** (časopisu)
- obavješćivanje – **elektronička pošta**
- **advanced search** mogućnosti i operatori





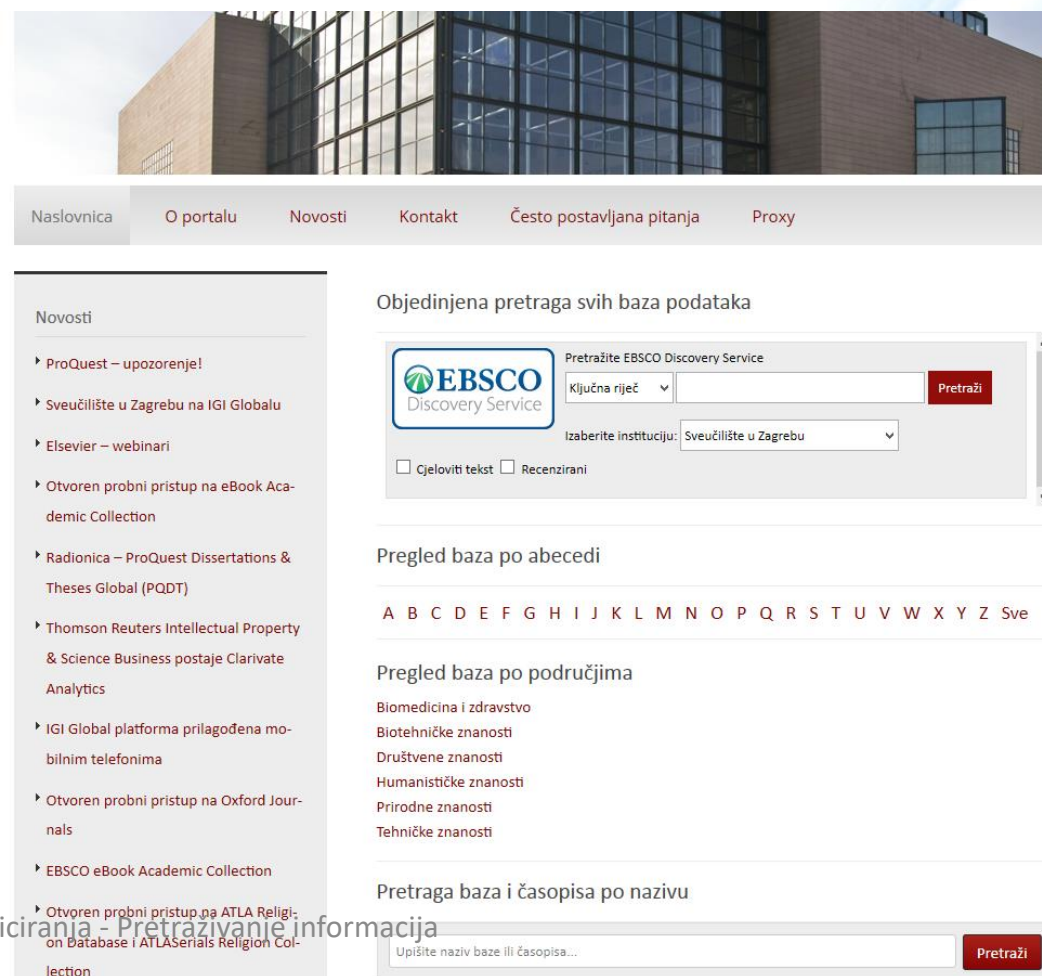
- greške u metapodacima zbog automatske ekstrakcije
  - godina je neki puta “izvučena” iz broja stranica, svezka, broja, broja telefona, broja faksa, poštanskog broja i drugih numeričkih podataka
  - fantomski autori, discipline, matične ustanove i nazivi časopisa
- niski kriteriji za indeksiranje



- online raspoložive knjige
- raspoloživost:
  - *Full view* – besplatno raspoloživo
  - *Snippet view* – samo djelići su raspoloživi
  - *No preview available* – nije raspoloživo
- mogućnosti pretrage slične kao Google (Scholar)



- pored Google scholar, Google **Books**
- on-line baze podataka koje indeksiraju znanstvene radove i časopise
- dostupne na <http://baze.nsk.hr/>



The screenshot shows the homepage of the Baze NSK website. At the top is a navigation bar with links: Naslovnica, O portalu, Novosti, Kontakt, Često postavljana pitanja, and Proxy. Below this is a large image of a modern building. The main content area is divided into two columns. The left column, titled 'Novosti', lists several news items with arrows pointing to them, including ProQuest – upozorenje!, Sveučilište u Zagrebu na IGI Globalu, Elsevier – webinar, and others. The right column features a search section titled 'Objedinjena pretraga svih baza podataka'. It includes the EBSCO Discovery Service logo and a search form with a 'Pretraži' button. Below the search form is a 'Pregled baza po abecedi' section with a list of letters from A to Z. Further down is a 'Pregled baza po područjima' section with a list of scientific fields. At the bottom is a 'Pretraga baza i časopisa po nazivu' section with a search bar and a 'Pretraži' button.

Naslovnica O portalu Novosti Kontakt Često postavljana pitanja Proxy

Novosti

- ▶ ProQuest – upozorenje!
- ▶ Sveučilište u Zagrebu na IGI Globalu
- ▶ Elsevier – webinar
- ▶ Otvoren probni pristup na eBook Academic Collection
- ▶ Radionica – ProQuest Dissertations & Theses Global (PQDT)
- ▶ Thomson Reuters Intellectual Property & Science Business postaje Clarivate Analytics
- ▶ IGI Global platforma prilagođena mobilnim telefonima
- ▶ Otvoren probni pristup na Oxford Journals
- ▶ EBSCO eBook Academic Collection
- ▶ Otvoren probni pristup na ATLA Religion Database i ATLASerials Religion Collection

Objedinjena pretraga svih baza podataka

Pretražite EBSCO Discovery Service

Ključna riječ  Pretraži

Izaberite instituciju: Sveučilište u Zagrebu

☐ Cjeloviti tekst ☐ Recenzirani

Pregled baza po abecedi

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z Sve

Pregled baza po područjima

- Biomedicina i zdravstvo
- Biotehničke znanosti
- Društvene znanosti
- Humanističke znanosti
- Prirodne znanosti
- Tehničke znanosti

Pretraga baza i časopisa po nazivu

Upišite naziv baze ili časopisa... Pretraži

- baze dostupne preko <http://baze.nsk.hr/> uključuju:
  - Web of Science i Current Contents
  - SCOPUS
  - IEEE Xplore - <http://ieeexplore.ieee.org/Xplore/>
  - ACM Digital Library - <http://dl.acm.org/>
- različiti kriteriji u izboru uključenih časopisa

- dodatno:
  - NSK - <http://www.nsk.hr/category/pretrazivanje/>
  - Library of Congress - <https://www.loc.gov/>
  - knjižnica FER-a -  
[http://www.fer.unizg.hr/knjiznica/online\\_katalog](http://www.fer.unizg.hr/knjiznica/online_katalog)



© Can Stock Photo - csp13611830

# A ŠTO KAD NAĐETE TRAŽENO?



# Spremite !!! Na organizirani način

The screenshot shows a Firefox browser window with the Zotero website open. The browser's address bar shows the URL <http://www.zotero.org/>. The website's header includes the Zotero logo and the tagline "The web now has a wrangler." Navigation links for "Home", "Groups", "People", "Support", and "Get Involved" are visible. A prominent red button labeled "download 1.0" is present, with the text "Latest version: 1.0.10 — maximum stability" below it. A search bar for "Search support" is also visible.

Below the website content, the Zotero extension interface is shown. It features a "My Library" sidebar with a list of folders: "Adaptation", "AdverseEffects", "BayesianStatistics", "ColoursfromtheSea", "Drugsfromthesea", and "FutureMedicalArticle". The main area displays a table of research sources with columns for "Creator", "Title", and "Journal Abbr".

Creator	Title	Journal Abbr
Allen	Sharing Scarcity: Bread Rationing and th...	
Altman	Better reporting of randomised controll...	BMJ
Altman	Endorsement of the CONSORT statemen...	BMJ
Altman et al.	The revised CONSORT statement for rep...	Ann Intern Med
Altman et al.	CONSORT statement requires closer exa...	BMJ
Ameer	Novel trial design: a report from the 19t...	J Clin Pharmacol
Armishaw and A...	Conotoxins as research tools and drug l...	Curr. Protein Pept. Sci
Balasubramania...	Standards of reporting of randomized c...	Ann. Surg
Batts et al.	Advancing Science through Conversatio...	
Becker and Terlau	Toxins from cone snails: properties, ap...	Appl. Microbiol. Biote...
Beckman	Are Bayes' days upon us? Statistical met...	J. Natl. Cancer Inst
Beckman	Are Bayes' Days Upon Us? Statistical Met...	
Begg et al.	Improving the quality of reporting of ra...	JAMA

On the right side of the extension interface, a "View Snapshot" panel is open, displaying details for a selected item: "Item Type: Journal Article", "Title: Conotoxins as research tools and drug leads", "Author: Armishaw, ...", "Author: Alewood, P...", "Abstract: The complex mixture of b...", "Publication: Current Protein & Peptide Science", "Volume: 6", "Issue: 3", "Pages: 221-240", and "Date: Jun 2005".

- **bibliografska** baza podataka
- desktop + browser **plug-in**
- **automatski** učitava podatke **s weba**
- može se dodati **full text**
- korisnik može **označiti** ključnim riječima, bilješkama
- može se organizirati u **hijerarhijske foldere**
- **sinkronizacija** s on-line
- **dijeljenje** s timom
- **alternativa:** Mendeley



- **Google**
  - upotrijebite napredno pretraživanje
  - nije sasvim objektivan
  - ima alternative
- postoje i agregatori - **meta tražilice**
- **ne vidi se sve** -> Deep Web
- za znanstvene i stručne informacije:
  - Google **Scholar**, Google **Books**
  - <http://baze.nsk.hr>
- kad nađete – **spremite**, organizirano - **Zotero**



# This is the way

- But I still haven't found what I'm looking for
  - [Google search documentation](#)
  - [Google calculator and conversions](#)
  - [DuckDuckGo Search Syntax](#)
  - [SciHub](#) (Wikipedia)
- What censorship?
  - [Censorship by Google](#) (Wikipedia)
  - [“Censored Planet: An Internet-wide, Longitudinal Censorship Observatory,”](#) (10.1145/3372297.3417883).
- Let me Google that for you
  - [Intelligent Assistants Have Poor Usability: A User Study of Alexa, Google Assistant, and Siri](#)
- Who needs my data anyway?
  - [Clues to Parkinson's and Alzheimer's From How You Use Your Computer](#) (2018)
  - [How Target figured out a teen girl was pregnant before her father did?](#) (2012)
  - [Social cooling](#) (2021)





**VJEKOM@ZESOI.FER.HR**

