



Универзитет „Св. Кирил и Методиј“ во Скопје
Педагошки факултет
„Св. Климент Охридски“ – Скопје

2. Веб-пребарувачи

Јасмина Јовановска

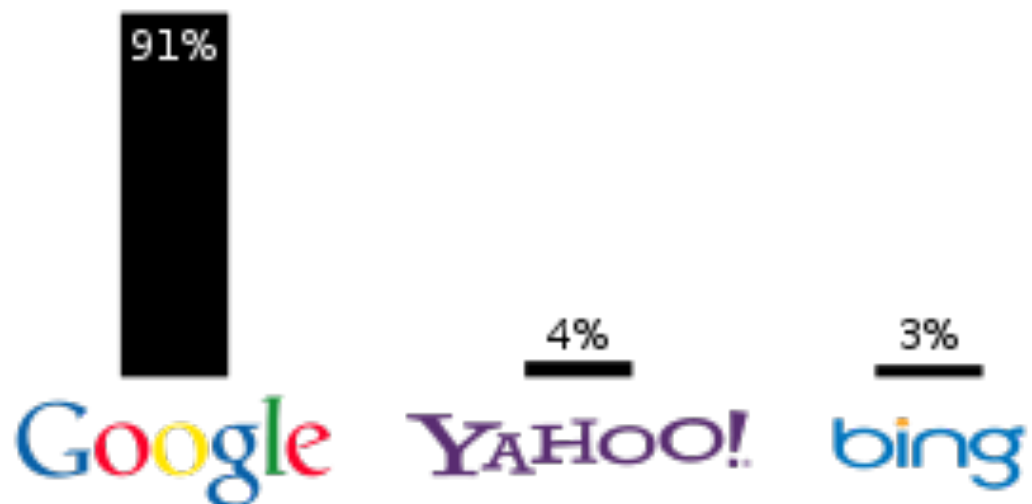
Педагошки факултет „Св. Климент Охридски“ - Скопје



2.1. Пребарување на информации на Веб

- **Веб-пребарувачите** се тип на системи за пребарување информации кои се развиени како одговор на потребата за наоѓање информации во огромни колекции со неструктурирани документи достапни на Веб (www).
- Постојат многу пребарувачи кои располагаат со различна:
 - содржина,
 - стратегии за доаѓање до одредени ресурси и
 - алатки кои им се располагање на корисниците.

2.1.1 Три најкористени веб-пребарувачи по процент на интернет корисници кон крајот на 2010 година



- На следниов линк може да ја погледнете листата на 15-те најпопуларни веб-пребарувачи:
 - <http://www.ebizmba.com/articles/search-engines>

2.2. Веб-пребарувачот Гугл



- **Google Inc.** или **Гугл** е американска корпорација специјализирана за интернет пребарувања. Помеѓу корисниците е најпозната по својот пребарувач Google, но покрај него нуди и многу други сервиси. Главно своите приходи ги остварува преку својот систем за рекламирање. Главното седиште на компанијата, познато под името Googleplex, се наоѓа во Калифорнија.
- Основачи на компанијата се тогашните студенти **Лери Пејџ** и **Сергеј Брин**. Целата приказна ја започнуваат неофицијално во јануари 1997 на **Универзитетот Стенфорд**, а официјално започнуваат со работа на 7 септември **1998** стартувајќи со својот интернет-пребарувач. Со своите сервиси успеваат својата основна дејност, пребарување и рекламирање, да ја прошират на други подрачја како што се: електронска пошта, онлајн мапи, прикажување на видео формати и слично.



2.2. Веб-пребарувачот Гугл

- <http://www.google.com/insidesearch/howsearchworks/thehistory/>
- Пребарувачот **Гугл** има огромен индекс од документи, многу е брз, поддржува Булови изрази, овозможува извлекување на пасуси. Индексираните страници, пресликани на серверите на Гугл за побрз пристап, вклучуваат не само HTML-документи, туку и PDF, Microsoft Office, итн.
- На кој начин функционира пребарувањето?
 - 1. Ползење (crawling) и индексирање (indexing),**
 - 2. Алгоритми,**
 - 3. Справување со спам.**

2.2.1. Ползење и индексирање



1. Ползење (crawling) и индексирање (indexing)

- Пребарувањето започнува со вебот кој се состои од над **60 трилиони веб-страници** и постојано се зголемува. Гугл „плови“ низ веб-просторот со ползење, односно ги следи врските од страница до страница.
- Ги сортира страниците според нивната содржина, како и според други фактори.
- Креира индекс кој е над **100 милиони гигабајти**.



2.2.2. Алгоритми

2. Алгоритми

- Имплементирани се повеќе програми и формули за да му се достават на корисникот најдобрите резултати.
- Со нив се извлекуваат заклучоци за подобро да се разбере прашалникот на корисникот. Врз основа на тие заклучоци се извлекуваат релевантни документи од индексот кои се рангираат врз основа на 200 фактори.
- Главна карактеристика на Гугл е неговиот **PageRank алгоритам** кој ги класифицира страниците врз основа на бројот на врски кои упатуваат кон нив.

PageRank



- При пребарување на документи релевантни на одреден прашалник (барање на корисникот), **PageRank алгоритмот** ги зема во предвид рангот на страниците, фреквенцијата и позицијата на термините од прашалникот, како и нивниот фонт и големите букви. Уште повеќе, страниците каде термините од прашалникот се појавуваат блиску еден до друг се сметаат за порелевантни.
- Уште една карактеристика која го олеснува пребарувањето е **извлекување на пасуси** од документите. Кога корисникот ќе постави прашалник, пребарувачот прикажува пасуси каде најголем дел од термините се појавуваат што е можно поблиску еден до друг.
- Целокупното пребарување на релевантните документи и враќањето на резултатите се случува за $1/8$ од секундата.

2.2.3. Справување со спам

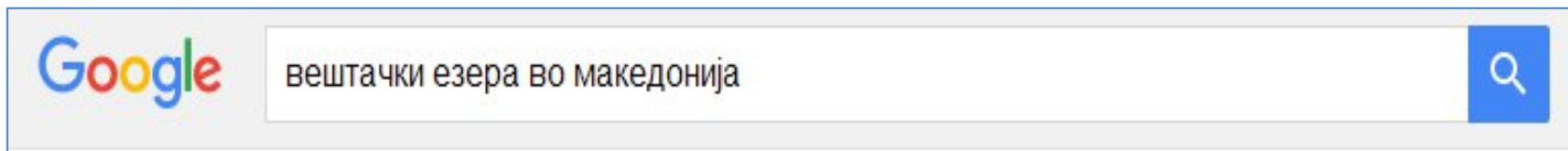


3. Справување со спам

- Гугл се справува со спамот 24 часа на ден за да се постигне подобра релевантност во пребарувањето.
- Најголем дел од отстранување на спамот се врши автоматски. Спорните документи се испитуваат рачно.

2.3. Трикови за пребарување на Гугл

- “” за пребарување на точна фраза

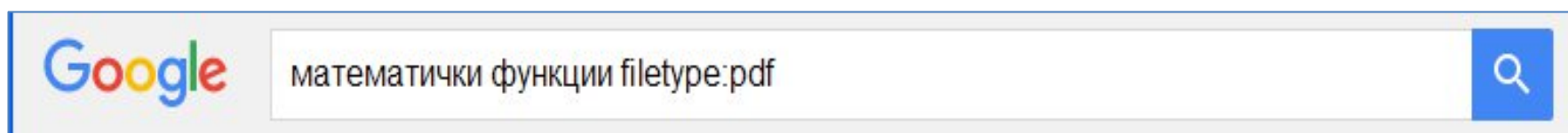


- Ако во полето за пребарување напишеме: вештачки езера во македонија, Гугл ќе прикаже резултати во кои зборовите “вештачки”, “езера” и “македонија” се појавуваат на различни места. Доколку истиот израз го ставиме во наводници, “вештачки езера во македонија”, тогаш ќе добиеме резултати во кои се појавува точно таа фраза со тој редослед на зборови. Ова е многу корисно за пребарување на книги, филмови или имиња на личности, организации, итн.

Трикови за пребарување на Гугл



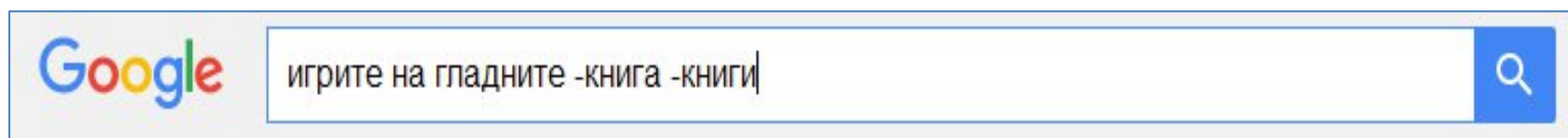
- **filetype:** за пребарување на одреден тип на фајл



- Може да пребаруваме тип на фајлови, на пример, да најдеме само PDF-фајлови со filetype:pdf или да најдеме Power Point презентации со filetype:pptx. Можеме да користиме и .avi, .docx, итн.

Трикови за пребарување на Гугл

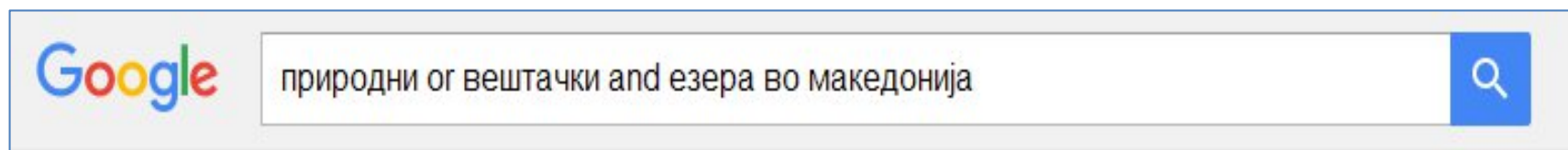
- Гугл ги користи и Буловите оператори: и, или, не.
- - за исклучување на зборови од пребарувањето (односно зборот да не се содржи во добиените резултати)



- Некогаш за одредени фрази постојат повеќе резултати. Да го разгледаме примерот со пребарување на фразата “игрите на гладните”. Бидејќи има и филмови и книги со тој наслов, доколку не сакаме да добиеме резултати за книгите можеме да користиме цртичка (минус) пред зборот кој сакаме да го исклучите од пребарувањето. На пример: игрите на гладните -книга -книги. Ова пребарување ќе даде резултати кои не ги содржат зборовите „книга“ или „книги“.

Трикови за пребарување на Гугл

- **OR** и **AND**: **ИЛИ** и **И**

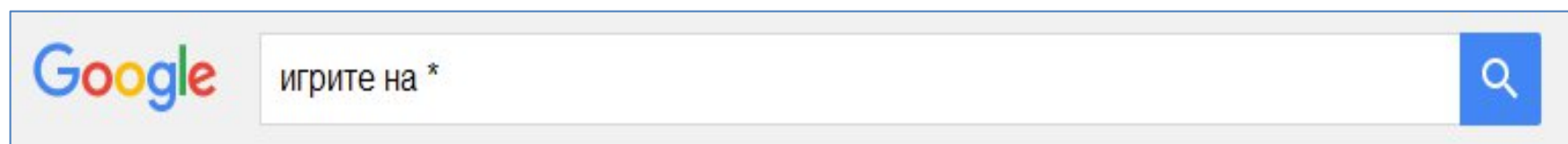


- При пребарувањето можеме да ги користиме и операторите OR и AND. На пример: доколку ги внесеме клучните зборови дадени погоре, тогаш специфицираме дека во добиените резултати треба да се појави зборот природни или зборот вештачки (или да се појават и двата) и задолжително да се појават зборовите езера во македонија.

Трикови за пребарување на Гугл



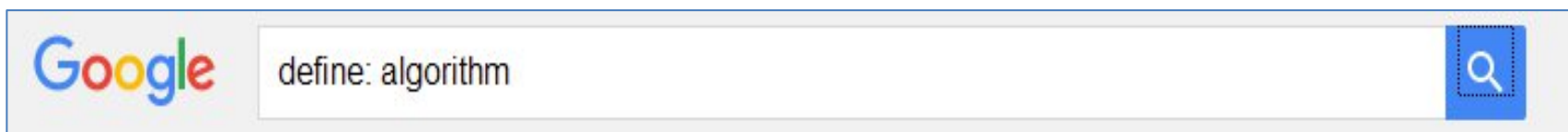
- * кога не знаеме збор во одредена фраза



- Доколку не ни текнува одреден збор од наслов на книга или филм, или било колку зборови од било каква фраза, можеме да искористиме * наместо зборот. На пример, можеме да пребаруваме игрите на * и ќе ни се прикажат резултати меѓу кои ќе се појави и бараната фраза, игрите на гладните.

Трикови за пребарување на Гугл

- **define:** да најдеме дефиниција и објаснување за одреден збор



- Доколку не ни е познато значењето на одреден збор, тогаш можеме да ја добиеме дефиницијата со помош на define:.

Трикови за пребарување на Гугл



- Постојат уште многу трикови за пребарување на Гугл
 - Употреба на симболот ~ за да во пребарувањето се вклучат и синоними на тој збор. На пример, ~problem во пребарувањето ќе ги вклучи и issue, dispute, итн.
 - **eur to mkd** - претворање на евра во денари,
 - **weather _** - за преглед на временската прогноза на одредена локација. На пример, weather skorje.
 - **movie:** - за да добиеме информации за филмови, итн.

THANK YOU!!!

