

Výlov novinek z rybníka zvaného Web Search

Vilém Sklenák – sklenak@vse.cz

Vysoká škola ekonomická, fakulta informatiky a statistiky,
katedra informačního a znalostního inženýrství

Inforum 2015, 26. 5. 2015

Obsah

- 1 Rybník Svět
- 2 Český rybníček
- 3 Co můžeme vidět (pod i nad vodou)

Je krásný na pohled...



Google na hladině

Dne 21. 4. 2015 nastal **Mobilgeddon...**

Google na hladině

Dne 21. 4. 2015 nastal **Mobilgeddon**...

- Ve výsledcích vyhledávání jsou upřednostňovány weby, které jsou „mobile friendly“.
- Týká se pochopitelně dotazů zadávaných z mobilních zařízení.
- Využívání mobilního Internetu v posledních letech strmě roste.
- Některé hodnocené vlastnosti: velikost písma, vzdálenost odkazů, šíře zobrazení atd.

<http://www.mobilmania.cz/clanky/google-zvyhodni-mobile-friendly-weby-mate-takov-sc-3-a-1330432/default.aspx>

Google pod hladinou

- Panda (znevýhodňování webů nižší kvality) i Penguin (penalizace webů s nekalými praktikami SEO) se dočkaly nových verzí.

Google pod hladinou

- Panda (znevýhodňování webů nižší kvality) i Penguin (penalizace webů s nekalými praktikami SEO) se dočkaly nových verzí.
- Pigeon (holub) – posilování pozic webů s lokálním cílením, zatím jen pro USA.

Google mezi jinými dravci

Kdo najde odvahu přijít s novým vyhledávačem?

Google mezi jinými dravci

Kdo najde odvahu přijít s novým vyhledávačem?

V posledních letech je to cesta lemovaná řadou pomníků
– Wikia Search, **Quaero**, **Theseus**, ...

Google mezi jinými dravci

Kdo najde odvahu přijít s novým vyhledávačem?

V posledních letech je to cesta lemovaná řadou pomníků
– Wikia Search, **Quaero**, **Theseus**, ... (EU nyní místo
technologického souboje vede s Googlem bitvy právní)

Google mezi jinými dravci

Kdo najde odvahu přijít s novým vyhledávačem?

V posledních letech je to cesta lemovaná řadou pomníků – Wikia Search, **Quaero**, **Theseus**, ... (EU nyní místo technologického souboje vede s Googlem bitvy právní)

Jediný úspěch na několikátý pokus je snad Bing (Microsoft) – **v USA podíl 20 %**, jinde však výrazně méně.

Google mezi jinými dravci

Kdo najde odvahu přijít s novým vyhledávačem?

V posledních letech je to cesta lemovaná řadou pomníků – Wikia Search, **Quaero**, **Theseus**, ... (EU nyní místo technologického souboje vede s Googlem bitvy právní)

Jediný úspěch na několikátý pokus je snad Bing (Microsoft) – **v USA podíl 20 %**, jinde však výrazně méně.

Novým vyzyvatelem by mohla být firma Apple – **již před časem koupila firmu Topsy, nyní má vlastního robota, prý pro Siri a Spotlight...**

Rozdíl mezi hlubinami a temnotami?

Rozdíl mezi hlubinami a temnotami?

Deep Web vs. Dark Web

Rozdíl mezi hlubinami a temnotami?

Deep Web vs. Dark Web

- Deep Web = hluboký/neviditelný web = obsah neindexovaný, ale běžně pomocí webu dostupný.
- Dark Web = úmyslně skrývaný obsah, běžnými prostředky nedostupný, mnohdy má ilegální charakter – viz třeba Tor.

Rozdíl mezi hlubinami a temnotami?

Deep Web vs. Dark Web

- Deep Web = hluboký/neviditelný web = obsah neindexovaný, ale běžně pomocí webu dostupný.
- Dark Web = úmyslně skrývaný obsah, běžnými prostředky nedostupný, mnohdy má ilegální charakter – viz třeba Tor.

Memex = memory extender

- 1945: Vannevar Bush použil při popisu principů hypertextu.
- 2015: DARPA oznamuje nový vyhledávač pro Dark Web.

Obsah

- 1 Rybník Svět
- 2 Český rybníček
- 3 Co můžeme vidět (pod i nad vodou)

Seznam mění tvář

Seznam.cz

Internet Firms Maps Slavic Zoology Directories Video

Vyhledat

Vybrat síťové Etniku.se

Dědili 25. května, svatka má Váňa

Asie / Letadla	Hry	Sportovní	Ceská republika	Ukolčení	
Svět / Házka	Letní	Národní	Wetor	V noci	Zima
Slavní	Mapy	Zobit	13 °C	9 °C	16 °C
On-line	Průběhy	Vše o			
Quozony	Recepty				
Praktický tip	SMS služby				

Novinky.cz

Čína potvrdila USA kvůli umělým ostrovům

Kuaj-gyanta amerického Spojených států P. P. Posádkou nář umělých ostrov, které Čína buduje v Jihočínském moři, postá...

Ušetřete v kině a na Divokém zvěřáku do televize a pak sledujte film

Řecko započalo s vydáváním vstupenek. Právě 6 restaurací

Ušetřete peníze na křesťanské svatoctví z kofe

Červen: 458. letice na nář letopisatelů 152. letice. Informace

Ušetřete peníze při výjezdu. Jeho majitel se odmění vyřizovat

Více zpráv o

Super.cz

Hvězda filmu Pearl Harbor je v Praze: Hollywoodského lešáka si s

Email.cz

Založit nový email

Jméno: seznam.cz

Město: PŘÍKAZ 36

Přidat se znovu na tomto počítači

Reklama

MLUVÍCÍ TLAKOMĚR SBM 52

JEN ZA 599,-

Tip

air/bank Získáte úročný běžný účet, který vydává, zařete si účet a Air Bank

Stáčí platit kartou

Další nabídky o

Firmy.cz

Seznam mění tvář



Seznam má rád jalapeño

- Boj s webovým spamem,
- snižování skóre u stránek používající agresivní SEO techniky,
- odhaduje se dopad na čtvrtinu hledání a změnu vedoucích pozic u 2% dotazů.



<http://seznam.seznamblog.cz/post/118765179956/seznam-cz-zkvalitnuje-sve-hledani>

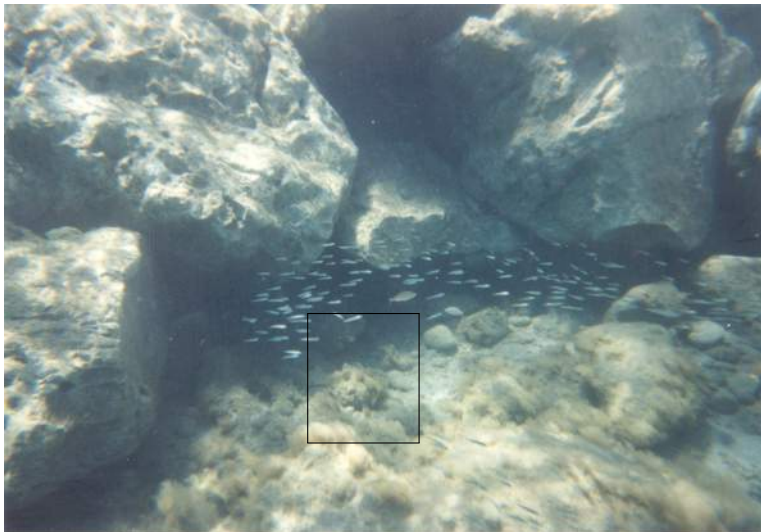
Obsah

- 1 Rybník Svět
- 2 Český rybníček
- 3 Co můžeme vidět (pod i nad vodou)

Co je to?



Co je to?



Co chudáci stroje?

- Princip „kouknu a vidím“ mnohdy nefunguje ani u člověka ;-)
- Indexování obsahu webu se nemůže vyhnout ani obrázkům.
- Z čeho lze vyjít:
 - ▶ název souboru,
 - ▶ hodnota atributu alt elementu img,
 - ▶ textový kontext podle umístění obrázku ve zdrojovém kódu,
 - ▶ případná metadata jako přímá součást obrázku, resp. souboru – EXIF (Exchangeable image file format).

Hledání obrázků podle Googlu

- Textové zadání dotazu – tedy víte, co chcete a dokážete to vyjádřit.
- Co když však nedokážete hledaný objekt výstižně pojmenovat?

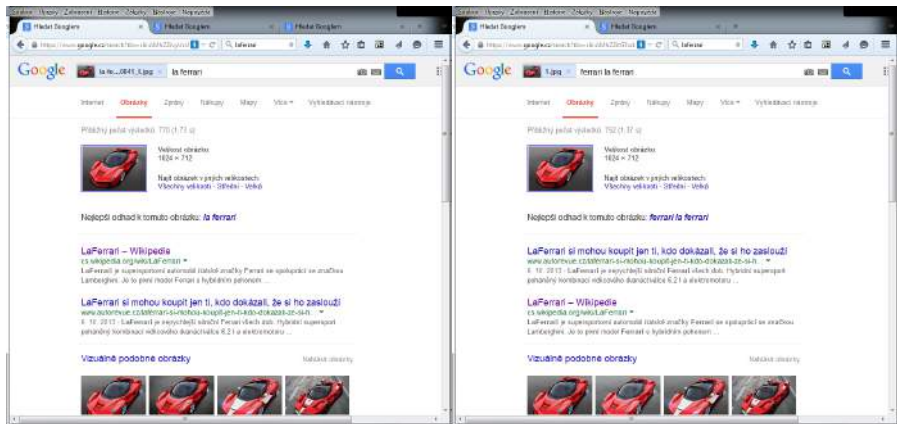
Hledání obrázků podle Googlu

- Textové zadání dotazu – tedy víte, co chcete a dokážete to vyjádřit.
- Co když však nedokážete hledaný objekt výstižně pojmenovat?
- Lze vyzkoušet **hledání podle obrazového vzoru**:



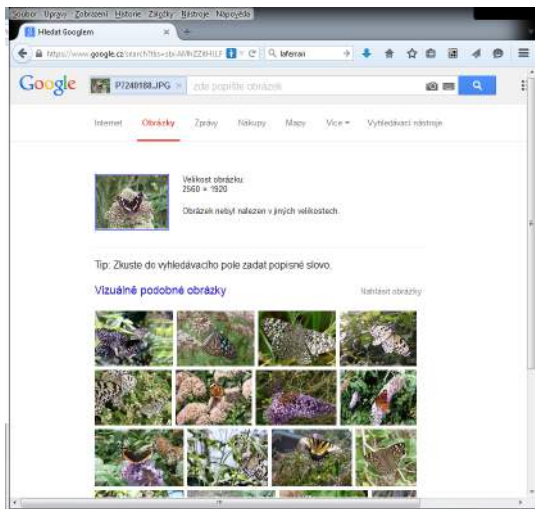
Hledání obrázků podle Googlu

Funguje to celkem pěkně, pokud se jedná o **obrázek pocházející z webových stránek**.



Hledání obrázků podle Googlu

Funguje to méně, pokud se jedná o **obrázek mimo web**.



Hledání obrázků – další vývoj

Stephen Wolfram – Image Identification Project

<https://www.imageidentify.com/>



Hledání obrázků – další vývoj

The screenshot shows a web browser window displaying the 'The Wolfram Language Image Identification Project' page. The browser's address bar shows the URL 'https://www.imagidentify.com/result/90831wvrat3'. The page title is 'The Wolfram Language Image Identification Project'. Below the title, there are social media sharing icons for Facebook, Twitter, Pinterest, and Email. A central image shows a red admiral butterfly on a cluster of small purple flowers. To the left of the image is a text input field containing 'ImageIdentify['. Below the image is a button with a magnifying glass icon and the text 'red admiral'. At the bottom, there is a feedback section with a checked checkbox 'Tell ImageIdentify how it did.' and four buttons: 'Great!', 'Could be better', 'Missed the point', and 'What the heck!'. The browser's search bar contains the text 'Hledat'.

Hledání obrázků – další vývoj

Google pracuje na **anotačním nástroji pro popis obsahu obrázku**. Využívá:

- techniky strojového učení
- a mohutný výpočetní výkon superpočítačů.

<http://googleresearch.blogspot.co.uk/2014/11/a-picture-is-worth-thousand-coherent.html>

Závěr

- Svět vyhledávání v posledních letech pronikl do každodenní praxe všech uživatelů Internetu. Internet pokrývá stále více domácností, navíc jej čím dále více uživatelů používá i na cestách.

Závěr

- Svět vyhledávání v posledních letech pronikl do každodenní praxe všech uživatelů Internetu. Internet pokrývá stále více domácností, navíc jej čím dále více uživatelů používá i na cestách.
- Vyhledávací služby se stávají čím dál více snáze použitelné.

Závěr

- Svět vyhledávání v posledních letech pronikl do každodenní praxe všech uživatelů Internetu. Internet pokrývá stále více domácností, navíc jej čím dále více uživatelů používá i na cestách.
- Vyhledávací služby se stávají čím dál více snáze použitelné.
- Odvrácená tvář: **zvyšuje se lenost mozku...**

Děkuji za pozornost.