

# Web 2.0, Search 2.0 – jak se to rýmuje?

Ing. Vilém Sklenák, CSc.

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

[sklenak@vse.cz](mailto:sklenak@vse.cz)

INFORUM 2008: 14. konference o profesionálních informačních zdrojích  
Praha, 28.–30. 5. 2008

## Abstrakt

Vliv principů Web 2.0 na vyhledávací stroje – na současné, vznik nových vyhledávacích strojů. Search 2.0 – co to znamená? Wikia – má šanci na úspěch? A co Quaero?

## 1 Úvodem

Tento příspěvek volně navazuje na příspěvky [8, 9, 10], které zazněly na jedné na konferenci Inforum 2007, jednak na předposledních konferencích AKP.

## 2 Co je Web 2.0

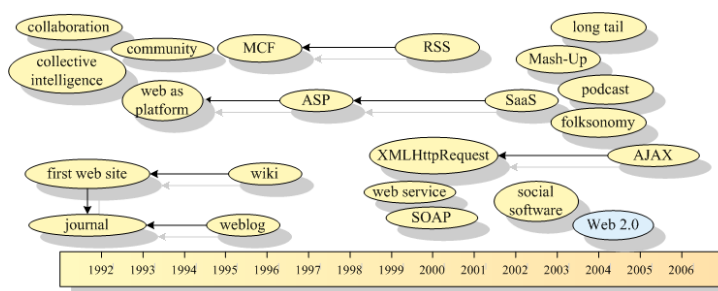
Pojem Web 2.0 byl poprvé použit v roce 2004 (Dale Dougherty), ale velkou měrou se na jeho popularizaci podílí Tim O'Reilly. Ten se pokusil na podzim 2005 ve svém článku [7] zformulovat jeho základní principy. Není to přístup založený na striktní definici, ale v článku se ukazují odlišnosti úspěšných internetových firem. Jde vlastně o seznam konceptů, metod a technologií. Web 2.0 je charakterizován jako posun od centralizovaného zpracování či služeb, což ukazuje tab. 1.

Technologie realizace:

- Tagy – označování obsahu uživateli, neboli folksonomy (jako protiklad k taxonomy).
- Sociální vztahy, příslušnost ke skupinám, souvisí s tagováním. Software, který bude poskytovat tyto služby, bude tím lepší, čím více bude uživatelů – to platí hlavně pro sociální systémy a folksonomy.

Web 1.0	Web 2.0
DoubleClick	Google AdSense
Ofoto	Flickr
Akamai	BitTorrent
mp3.com	Napster
Britannica	Online Wikipedia
personal websites	Blogging
evite	upcoming.org and EVDB
domain name speculation	search engine optimization
page views	cost per click
screen scraping	web services
publishing	Participation
content management systems	Wikis
directories (taxonomy)	tagging („folksonomy“)
stickiness	Syndication

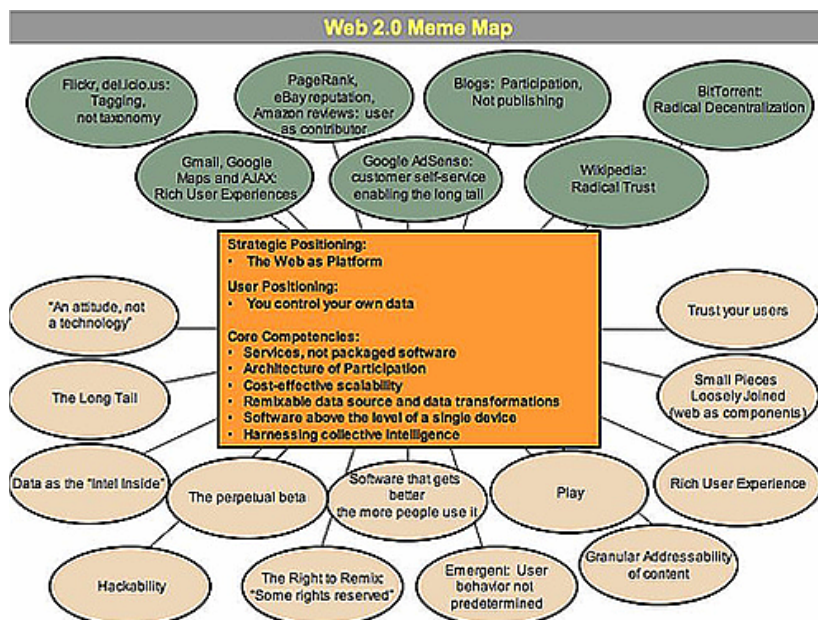
Tabulka 1: Odlišnosti „starého“ a „nového“ webu



Obrázek 1: Vývoj související s pojmem Web 2.0

- RSS – syndikace obsahu. Uživatel nebude muset hledat, co je nového, ale informace k němu potečou přes zvolené kanály.
- Long tail – blogy, RSS, Google adsense – samoobsluha pro zákazníky, hodně malých zdrojů.
- Blogy a Wikipedia – uživatel jako přispěvatel a tvůrce obsahu. Blogy jako zdroj obsahu spolu s RSS znamenají rozšíření obsahu, uživatelé se na něm podílejí – píšou zprávy a úvahy, které byly dříve jen v centrálních médiích.
- Programové nadstavby pro zpracování dat: AJAX – technologie, která umožňuje asynchronní přenos dat, aplikace nad obsahem databází a webů, webová rozhraní (API) pro přístup nadstavbových aplikací třetích stran.

- Vznikne něco jako webové služby, které dokáží propojit data, např. RSS a další webové služby pro přenos dat (SOAP – simple object access protocol) a REST – (Representational data transfer, přenos přes xml formát) – používá např. Amazon.com.



Obrázek 2: Představa Webu 2.0

### 3 Search 2.0

Vlna zájmu o Web 2.0 měla a má vliv na řadu služeb, které se bez webové platformy neobejdou a které tak mohou z principů Webu 2.0 leccos vytěžit při dalším zkvalitňování. U vyhledávacích služeb tomu nebylo jinak. V diskusích či různých materiálech se začínají objevovat *Search 2.0* či *Search 3.0* a výše. Co to však znamená?

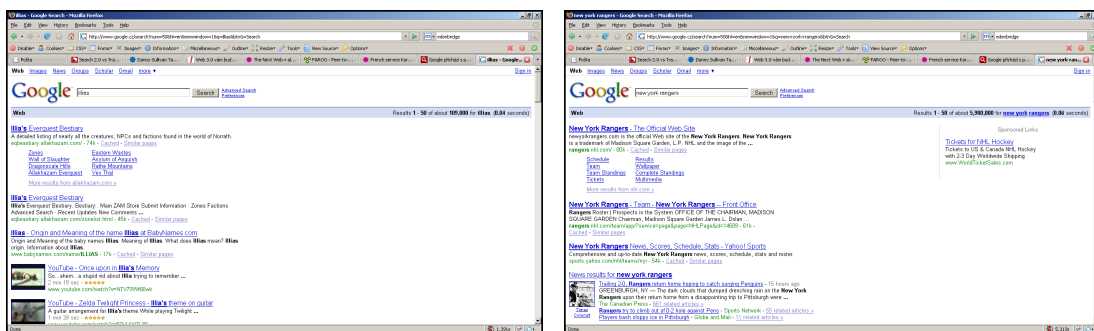
U různých autorů lze totiž vysledovat různé pohledy. Například Ebrahim Ezzy v [4] při vymezení obsahu *Search 2.0* vychází z dělení vyhledávacích strojů do generací:

- 1. generace je typická řazením výsledků podle obsahu dokumentů – viz třeba původní AltaVista;
- 2. generace používá pro řazení výsledky analýzy struktury odkazů mezi dokumenty na webu – viz Google;

- 3. generace = Search 2.0. Má jít o nové vyhledávací technologie, které jsou vhodné ke zkombinování možností současných vyhledávacích strojů s novým a zdokonalenými způsoby určování relevance, čímž se myslí brát ohled na různé preference uživatelů, jejich spolupráci, jejich zkušenosti a jiné specializované možnosti, které vcelku přispějí k větší efektivitě/vytěžitelnosti nalezených informací.

V [4, 5] jsou rovněž uvedeny příklady řady vyhledávacích služeb „nového“ typu stejně jako na stránce [6] – *FlickrStorm*, *FundooWeb*, *Similicio.us*, *Kartoo*, *Kw-Map*, *MnemoMap*, *Quintura*, *Liveplasma*, ... Nutno poznamenat, že zcela po právu tyto nové vyhledávací služby nemají ambice přímo konkurovat zavedeným vyhledávacím službám, ale chtějí ukázat nové možnosti, které mnozí uživatelé očekávají a které mohou vést k zajímavému synergickému efektu.

Trochu odlišný pohled na obsah *Search x.0* je uveden v [11], kdy číslování „verzí“ se odvíjí od generací naznačených v předposledním odstavci. A proto jako Search 2.0 jsou chápány nástroje, které pro řazení výsledků vyhledávání těží z analýzy odkazů na webu. Za *Search 3.0* jsou označovány přístupy, které kombinují různé typy vyhledávání do jednoho výsledku. Příkladem je tzv. *univerzální vyhledávání*, které uplatňuje Google od roku 2007 a které kombinuje jak hledání na webu, tak hledání ve zpravodajství, ve videosekvencích, obrázcích apod.



Obrázek 3: Google – univerzální vyhledávání

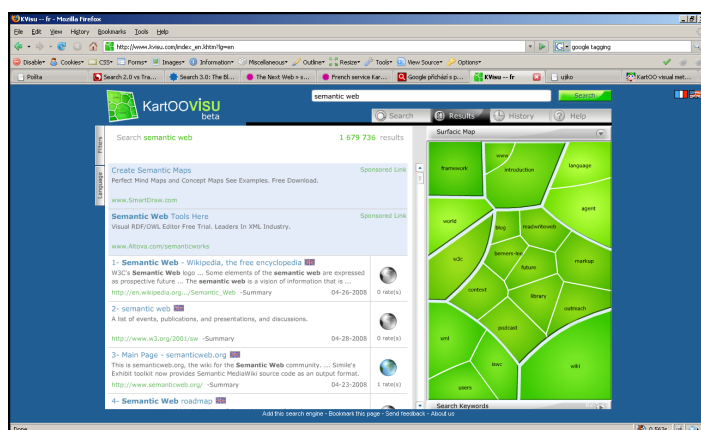
A teprve *Search 4.0* pak má využívat sociální sítě uživatelů. . .

Ono číslování asi není úplně to podstatné – závěr se přímo nabízí, totiž že významnou roli hrají uživatelé a jejich poptávka. Dobře o tom svědčí také blog *The Next Web* [1], který funguje od roku 2006 a který je spojen s každoroční stejnojmennou konferencí. V textech na tomto blogu, ať už se jedná o komentáře, nebo vystoupení z několika ročníků konferencí, žádné *Search x.0* nenajdete. Otázky trendů ve vyhledávání či vlivu moderních webových technologií na vyhledávací služby však diskutovány samozřejmě jsou. I zde se projevuje jako výrazný trend socializace vyhledávání. Například v poslední době zde byly představeny služby *Eeggi*, *Faroo*, *Stumpedia*, ...

Obrázky 4-7 zachycují některé netradiční přístupy vyhledávacích služeb.



Obrázek 4: Kartoo – metavyhledávací stroj



Obrázek 5: KartooVisu – prezentace výsledků v podobě mapy

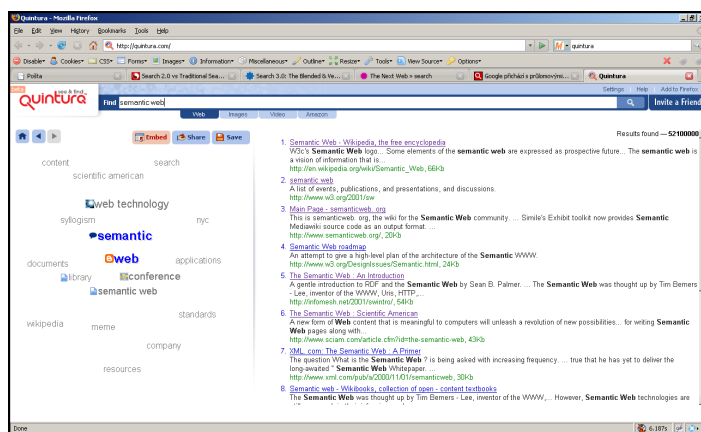
## 4 Wikia Search

Zdá se, že firma Google představuje jak pro již existující konkurenci, tak pro konkurenci potenciální, minimálně výzvu, ve většině případů však spíše určitou „červenou muletou“. Jedním z příkladů z poslední doby je vyhledávací stroj *Wikia Search* [13], jehož duchovním otcem je Jimmy Wales, jinak také zakladatel tolik populární Wikipedie.

Východiskem fungování tohoto stroje má být opět komunitní přístup, tedy podobně jako u Wikipedie a v duchu rysů Webu 2.0. První veřejná verze byla zpřístupněna zhruba rok po zveřejnění záměru. Každý zájemce si ji může otestovat



Obrázek 6: Ujiko – personalizovaná nadstavba nad vyhledáváním pomocí Yahoo

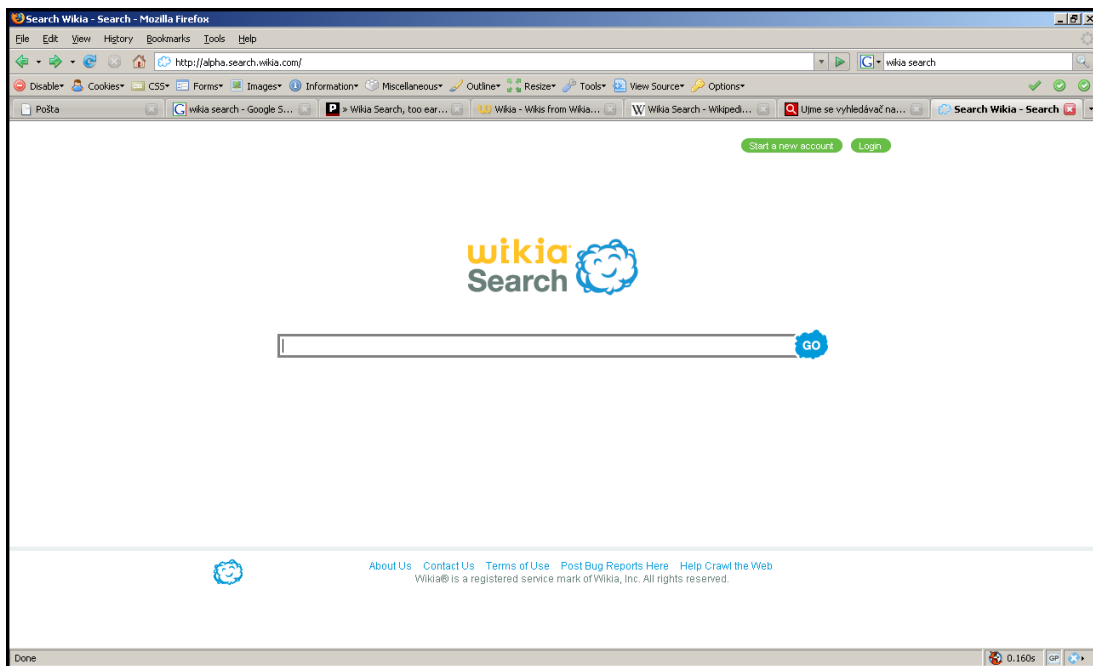


Obrázek 7: Quintura – výsledky vyhledávání jako tagclouds

na adrese <http://alpha.search.wikia.com/> – jak již samotné URL napovídá, jedná se stále o testovací verzi ve fázi alfa testování. Za poměrně krátkou dobu mezi záměrem a zpřístupněním vzniklo poměrně nadějně očekávané poplatné popularity Wikipedie. Velmi záhy se také objevily glosy od komentátorů, bloggerů aj., které byly plně zklámaní a nenaplněných šancích apod.

Je dobré si však uvědomit, co je uvedeno při vstupu na web *Wikia Search*, tj. <http://search.wikia.com/>:

*Search Wikia is still very much under development, and we are aware that the quality of the search results is low. What you see here is our first alpha release. ... Wikia's search engine concept is that of trusted user feedback from a community of users acting together in an open, transparent, public way. Of course, before we start, we have no user feedback data. So the results are pretty bad. But we expect them to*



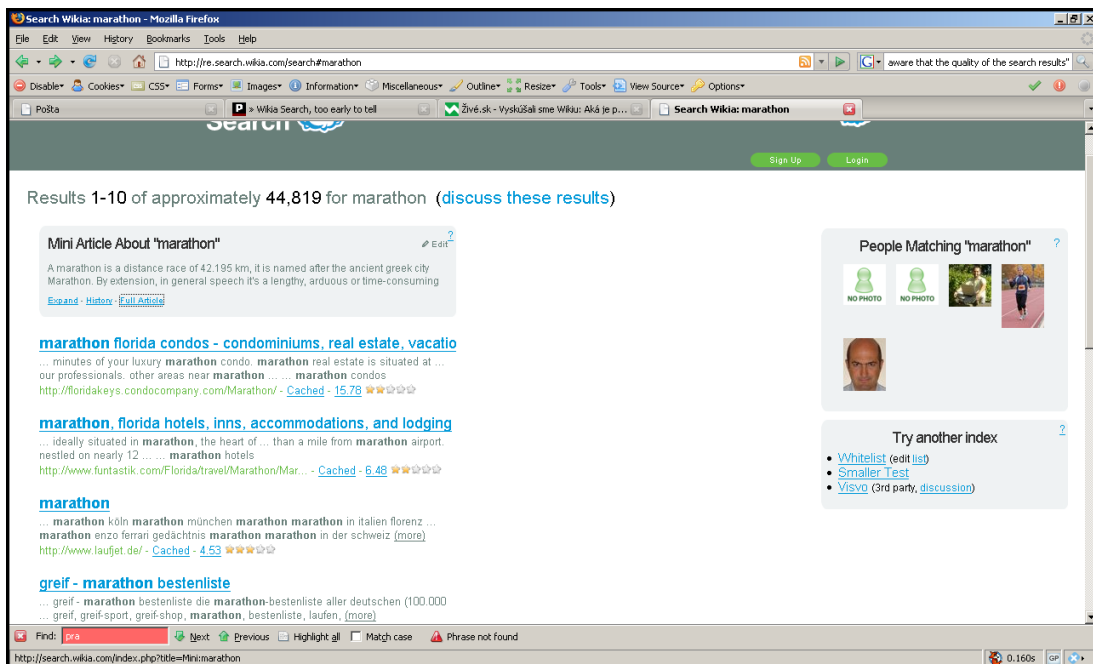
Obrázek 8: Wikia Search

*improve rapidly in coming weeks, so please bookmark the site and return often.*

Jinými slovy je tento nástroj ve stádiu růstu, nemá kvalitní databázi a nemá ani tolik potřebnou komunitu uživatelů – to vše by mělo přijít časem. Vyhledávání funguje – aspoň tak lze soudit podle obdržených výsledků. Zatím chybí nápověda – jaké formulace dotazů jsou možné, zda jsou možné logické operátory apod. Víceslovné dotazy jsou vyhodnocovány pravděpodobně jako implicitní AND, pořadí slov v dotazu nemá vliv na pořadí výsledků, uvozovky pravděpodobně fungují jako kontextový operátor, ...

Prezentace výsledků obsahuje na první pohled některé odlišnosti:

1. Možnost hodnocení – již zavedený princip „hvězdičkování“. Přidělování hvězdiček možné sice je, ale zatím se tento rys stále testuje, a proto nejsou hodnocení nikam ukládána.
2. Miničlánky – krátká, shrnující informace o hledaném pojmu, která se objevuje na počátku výsledků (viz obr. 9). Tento rys funguje podle očekávání u jednoslovných dotazů, u víceslovných bych očekával miničlánky pro jednotlivá slova včetně případné fráze. Tímto způsobem to zatím nefunguje, ačkoliv jsem zkoušel dotazy ze slov, u kterých u jednoslovných dotazů miničlánek byl.



Obrázek 9: Výsledky vyhledávání Wikia Search

Na hodnocení služby Wikia Search je stále brzo – nelze ji ani zavrhnout, ani chválit. Je nutno ji dát čas, aby dozrála. Zdá se, že start byl možná trochu uspěchaný, ani po několika měsících některé klíčové funkce vlastně nefungují. Takže lze jen doufat, že se jí podaří vybudovat tolik potřebnou komunitu uživatelů, kteří se nenechají odradit tímto počátečním tápáním.

## 5 Quaero, Theseus, ...

Quaero [12] je projekt „evropského“ vyhledávacího stroje, jeho záměr byl vyhlášen na jaře 2005, některé práce pak započaly o rok později. Název je převzat z latiny a znamená *hledám*. Projekt je v současnosti především záležitostí francouzských firem (více než 20), ale zpočátku to vypadalo na francouzsko-německou spolupráci, ze které však německá strana odstoupila koncem roku 2006, když současně vyhlásila projekt vlastní – Theseus.

Hlavní doménou Quaera má být vyhledávání v multimediálním prostředí. Vyhledávací nástroj bude využívat techniky pro rozpoznání, indexování a automatizovaný překlad audiovizuálních dokumentů. Zde je míněno prohledávání jednak obrazových záznamů na základě rozeznávání tvarů a barev, nikoliv pouze vyhledávání v metadatech, která soubory popisují; jednak vyhledávání ve zvukových záznamech a jejich převod do textové podoby, včetně překladu.



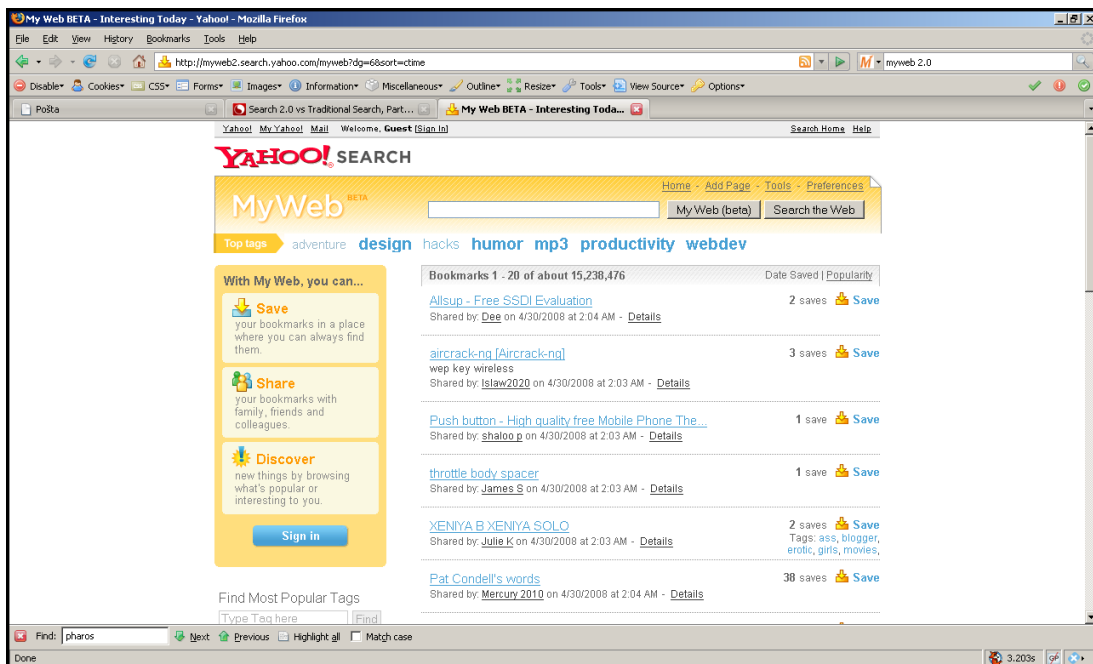
Projekt Quaero byl podrobně představen na tiskové konferenci dne 17. 3. 2008, tedy zhruba týden poté, co získal definitivní potvrzení na finanční dotaci – celkový rozpočet je více než 200 miliónů EUR (99 mil. poskytne francouzská vláda, zbytek je vklad partnerů). Více informací je o projektu na projektovém webu: <http://www.quaero.eu>. Zatímco Quaero má ambice být konkurencí pro Google (proto někteří komentátoři používají přirovnání souboji Airbus vs. Boeing), Theseus jde jinou cestou [3]. Cílem projektu Theseus jsou vyhledávací technologie, které budou moci být začleněny do jiných vyhledávacích strojů a nástrojů. Důraz je kladen na obraz a video. Klíčovou roli mají sehrát sémantické technologie vyhledávání, které mají posloužit při analýze kontextu vyhledávané informace. První prototyp se předpokládá v roce 2011.

## 6 Závěrem

Svět vyhledávacích služeb není strnulý – musí reagovat jednak na potřeby uživatelů, jednak na nové technologické možnosti. Ty umožňují nabídnout uživatelům nové funkce, o kterých dosud uživatelé třeba ani nevěděli, že by je mohli chtít. V souvislosti s vlnou zájmu a popularizace přístupů označovaných jako Web 2.0 vznikla také řada námětů, které by mohly ukázat určitý směr vývoje i pro rozhodující „hráče“ na trhu, jako jsou Google, Microsoft či Yahoo. Je také pravda, že u některých z nich se už objevily snahy začlenit sociální vyhledávání do svého portfolia. Například Yahoo provozuje službu *MyWeb 2.0* (<http://myweb2.search.yahoo.com/>). S jistou nadsázkou lze říci, že budoucí možnosti i velkých vyhledávacích služeb mají trochu v rukou i sami uživatelé.

## Reference

- [1] The next web:. URL <http://thenextweb.org>.
- [2] The Quaero project – new european search technology. 2007. URL <http://www.pandia.com/sew/553-quaero-2.html>.
- [3] The Theseus image and video search project. 2007. URL <http://www.pandia.com/sew/570-theseus.html>.
- [4] Ezzy, Ebrahim. Search 2.0 vs traditional search. ReadWriteWeb, 2006. URL [http://www.readwriteweb.com/archives/search\\_20\\_vs\\_tr.php](http://www.readwriteweb.com/archives/search_20_vs_tr.php).
- [5] Ezzy, Ebrahim. Search 2.0 vs traditional search, part 2. ReadWriteWeb, 2006. URL [http://www.readwriteweb.com/archives/search\\_20\\_vs\\_tr.php](http://www.readwriteweb.com/archives/search_20_vs_tr.php).
- [6] OEDb. Top 25 web 2.0 search engines. 2007. URL <http://oedb.org/library/features/top-25-web20-search-engines>.
- [7] O'Reilly, Tim. Whar is web 2.0. design patterns and business models for the next generation of software. 2005. URL <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>.



Obrázek 10: Yahoo – MyWeb 2.0

- [8] Sklenák, Vilém. Vyhledávací nástroje v prostředí internetu – co bude dál? In Tkačíková, Daniela; Ramajzlová, Barbora (eds.) *Automatizace knihovnických procesů – 9*. ČVUT – Výpočetní a informační centrum, 2003, s. 22–30. URL [http://knihovny.cvut.cz/akp2003/sbornik/03\\_sklenak.pdf](http://knihovny.cvut.cz/akp2003/sbornik/03_sklenak.pdf).
- [9] Sklenák, Vilém. Vyhledávání informací v prostředí webu – mírný pokrok v mezích zákona. In Tkačíková, Daniela; Ramajzlová, Barbora (eds.) *Automatizace knihovnických procesů – 10*. ČVUT – Výpočetní a informační centrum, 2005, s. 74–79. URL <http://www.akvs.cz/akp-2005/10-sklenak.pdf>.
- [10] Sklenák, Vilém. Web 2.0 vs. sémantický web. In *Inforum 2007*. Albertina icome Praha, 2007. URL <http://www.inforum.cz/pdf/2007/sklenak-vilem.pdf>.
- [11] Sulliovan, Danny. Search 3.0: The blended & vertical search revolution. Searchengineland, 2007. URL <http://searchengineland.com/071127-091128.php>.
- [12] Wikipedia. Quaero. 2008. URL <http://en.wikipedia.org/wiki/Quaero>.
- [13] Wikipedia. Wikia search. 2008. URL [http://en.wikipedia.org/wiki/Wikia\\_search](http://en.wikipedia.org/wiki/Wikia_search).