

Web 2.0 vs. sémantický web

Ing. Vilém Sklenák, CSc.

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

sklenak@vse.cz

INFORUM 2007: 13. konference o profesionálních informačních zdrojích
Praha, 22.–24. 5. 2007

Abstrakt

V poslední době jsme svědky stále častějšího používání pojmu *web* s různými pří-
vlastky či označením verzí – sémantický web, Web 2.0, Web 3.0, Cílem příspěvku
je ukázat odlišnosti i spojitosti a nastínit i předpokládaný vývoj.

1 Úvodem

Tento příspěvek volně navazuje na příspěvek [14], který zazněl na konferenci Inforum 2003. V tehdejší příspěvku byly shrnuty základní principy sémantického webu. Jako počátek vývoje se obvykle uvádí článek Tima Bernerse-Lee [2] obsahující základní tezi – „*sémantický web je rozšířením současného webu, jež datům přiřazuje přesný význam, díky kterému bude možná kooperace jak lidí, tak softwaru*“. Závěr příspěvku před čtyřmi lety byl otevřený – „*sémantický web má za sebou již několikaletý výzkum, ale přesto se nachází v počítačném údobí svého vývoje. Mnoho bylo uděláno v oblasti jazyků, reprezentace znalostí, ontologií. Další krok, který bude následovat, je vývoj aktivních komponentů, které budou nabízet uživatelům inteligentní služby*“. Nyní se zdá, že další vývoj je potvrzením. Začalo se hovořit o technologických zdokonaleních webu – začaly se prosazovat pojmy jako Web 2.0, Web 3.0 apod. To ostatně dokresluje i obr. 1 – strmě rostoucí čára od r. 2005 reprezentuje Web 2.0, čára na vzestupu od konce r. 2006 reprezentuje Web 3.0 a napříč se táhne čára konstatního zájmu o sémantický web.

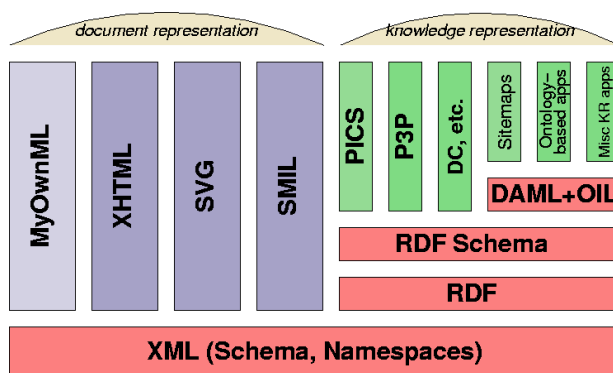
2 Co je sémantický web

Jak plyne z již citovaného článku [2] sémantický web není nějaký nový web, ale jde o rozšíření konceptu a doplnění dat toho stávajícího. Doplnění o metadata, která by měla popisovat sémantické informace webových zdrojů a která by měla být zápsána pomocí strojově srozumitelných jazyků. Součástí metadat by také byla použitá slovní zásoba a soubor vztahů mezi jednotlivými pojmy.



Obrázek 1: GoogleTrends – semantic web, web 2.0 a web 3.0

Na webu je však téměř nemožné prosadit jednotný jazyk a vymezit jakousi jednotnou slovní zásobu. Plyne to jednak z principu decentralizovanosti samotného webu, jednak z povahy zpřístupňovaných informací – jde vlastně o všechny oblasti znalostí.



Obrázek 2: Jazyky sémantického webu

O to se však sémantický web nesnaží. Jeho myšlenka spočívá především v nabídce takového flexibilního a otevřeného datového modelu a odpovídajících datových jazyků tak, aby vyhovoval nekonečně varietě webu.

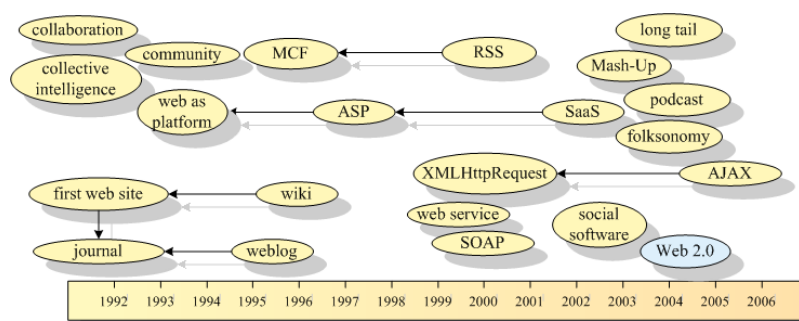
3 Co je Web 2.0

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 [12] 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:

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

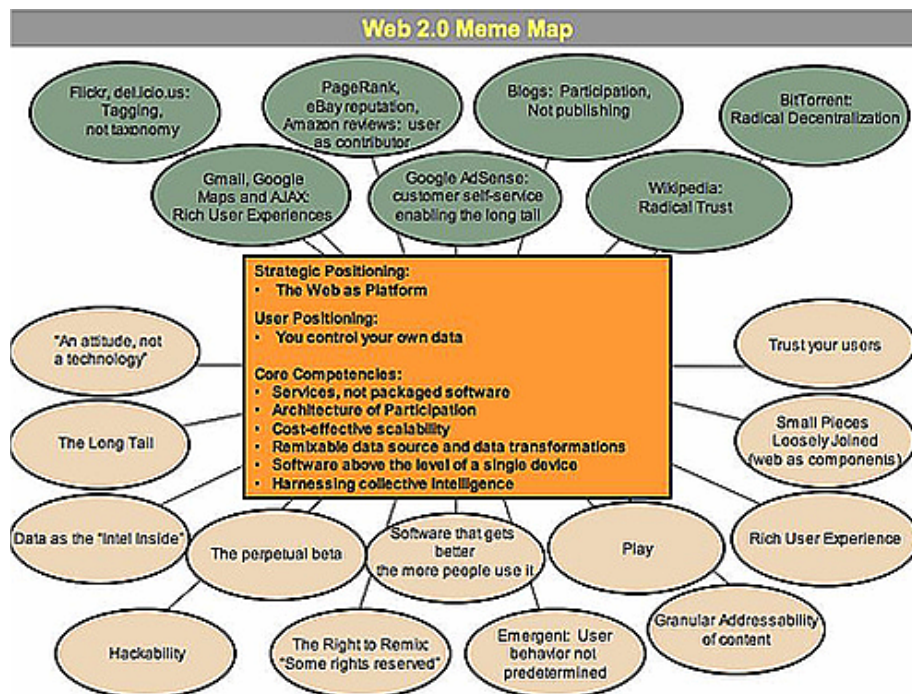
Tabulka 1: Odlíšnosti „starého“ a „nového“ webu



Obrázek 3: Vývoj související s pojmem Web 2.0

- 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.
- 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 4: Představa Webu 2.0

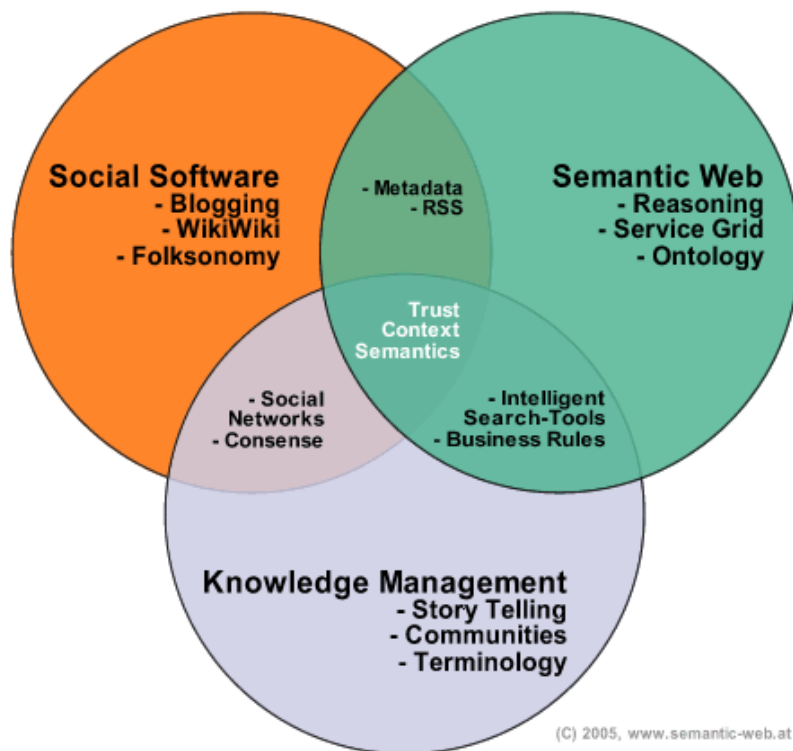
4 Web 2.0 vs. sémantický web

Zákonitě velmi krátce poté, co se začalo diskutovat o konceptu Webu 2.0, vznikla otázka vztahu mezi Webem 2.0 a sémantickým webem, jehož myšlenky vznikly o poznání dříve. Spektrum názorů bylo a stále je velmi široké. Na jedné straně byly názory, že myšlenka sémantického webu je již vyčpělá a že Web 2.0 je to pravé, což se obvykle vyjadřovalo stručně – „*Web 2.0 is killing Semantic Web*“. Na straně druhé jiní poukazovali, že Web 2.0 je pouhý marketingový tah firem z oblasti informačních technologií. Kde je pravda?

Ta bude někde uprostřed. Jde totiž o to, že Web 2.0 a sémantický web mají mnoho společného – obojí se snaží o sdílení a znovupoužitelnost dat v rámci různých komunit, pro obojí je typická snaha o integraci dat. Realističtěji proto vyznívá pohled, že se oba přístupy vhodně doplňují – Web 2.0 představuje platformu/technologie a sémantický web pak význam. Tento názor ostatně zastává také Tim Berners-Lee [1], který říká, že se nejedná o alternativy (doslova používá příměr „chalk and cheese“, čili „nebe a dudy“) a že si naopak mají co nabídnout:

- Web 2.0 „neumí“ integraci mezi aplikacemi,
- sémantický web „neumí“ atraktivní rozhraní,

a proto je podnětné uvažovat o provázání.



Obrázek 5: Sématický web, social software, Web 2.0 – souvislosti

Schematicky jsou některé souvislosti Webu 2.0 a sémantického webu ukázány na obr. 5. Co se rýsuje jako možná kombinace idejí Webu 2.0 a sémantického webu:

- **Sémantické wiki** – zajímavá výzkumná oblast, jejímž cílem je nalezení způsobů, jak sémanticky anotovat obsah wiki. Díky tomu by byl možný formální popis obsahu s využitím takových jazyků jako RDF a OWL.

- **Sémantické blogování** – blogy mohou znamenat více než jednoduchý nástroj pro publikování osobních názorů. Budeme-li uvažovat schopnost generovat strojově čitelné RSS, pak není daleko k distribuci a agregaci podobných informací z mnoha různých zdrojů.
- **Sémantický desktop** – lze uvažovat, že kombinací idejí sémantického webu a Webu 2.0 s tradičními desktopovými aplikacemi a jejich daty (dokumenty, e-maily, fotky) může vzniknout personalizovanější styl práce. Jde hlavně o to, jak umožnit uživateli snazší orientaci ve svých datech (někdy se používá poněkud nelichotivé označení „personal data swamps“), nebo aby se v nich vůbec vyznal.
- **Práce s ontologiemi a folksonomiemi** – existují výzkumy, jejichž cílem je zachování flexibility volného tagování pro anotování, ale pro prohlížení nebo vyhledávání by se používala ontologie.

4.1 Web 2.0 a vyhledávací stroje

Web 2.0 přichází s novými technologiemi, s novými vztahy a možnostmi publikování, a proto se postupně objevují alternativní vyhledávací stroje, které jsou založeny na technologiích Webu 2.0. Za zajímavé lze považovat:

- **Kartoo**¹ – shlukování výsledků v podobě grafu s označenými hranami,
- **Quintura**² – shlukování s využitím principu „tagclouds“,
- **KwMap**³ – po vyhledání nabízí relevantní doplňková klíčová slova,
- **Whonu**⁴ – nabízí kontextové rozšíření dotazu již během jeho psaní,

Podrobný a komentovaný přehled je uveden v [11] nebo v [7].

4.2 Sémantický web a vyhledávací stroje

Vyhledávací stroje jak pro sémantický web, tak pro „konvenční“ web jsou si na nejvyšší úrovni velmi podobné: získávání a vytěžování dokumentů; zpracování dotazů od uživatelů a agentů; řazení výsledků vyhledávání; archivace dokumentů; nabízet rozhraní jak pro uživatele, tak aplikační rozhraní pro software. Obrázek 6 ukazuje základní architekturu vyhledávacího stroje Swoogle, který je specificky určen pro vyhledávání v prostředí sémantického webu.

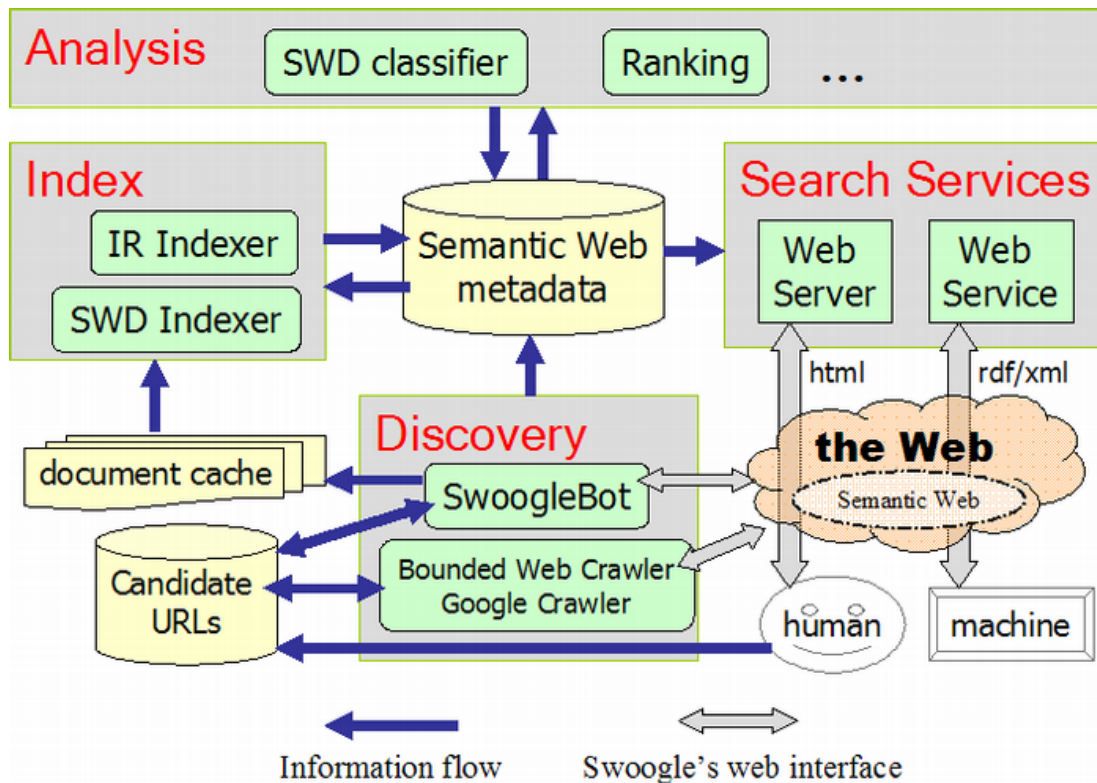
Odlišnosti vyhledávání v sémantickém webu mají vícero příčin:

¹<http://www.kartoo.com>

²<http://www.quintura.com>

³<http://www.kwmap.com>

⁴<http://www.whonu.com>

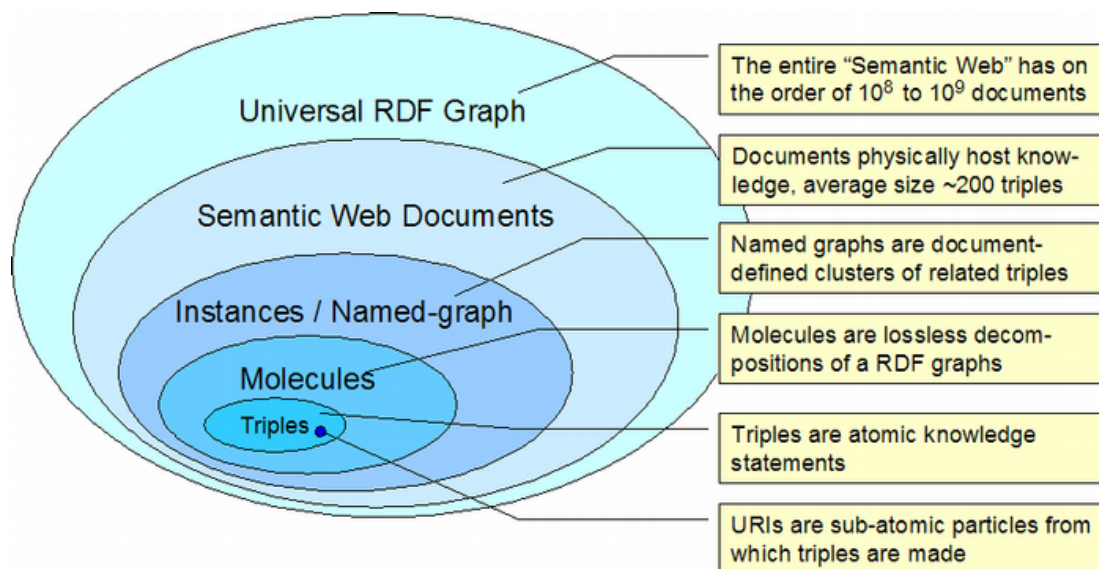


Obrázek 6: Architektura vyhledávacího stroje swoogle

1. Posláním sémantického webu je mj. strojové zpracování obsahu, proto použité značkování či anotování může dobře posloužit při konstrukci agentské vrstvy mezi uživatelem a vyhledávacím strojem.
2. Znalosti, které jsou zachyceny např. pomocí RDF, se liší jak od převážně nestrukturovaných dokumentů běžného webu, tak od dobře strukturovaných informací v databázích (např. relačního typu). Takto semistrukturované informace vyžadují kombinaci technik pro efektivní indexování a vyhledávání.
3. Jeden dokument sémantického webu může být směsicí konkrétních faktů, tříd a definic vlastností, logických omezení a metadat. Plné porozumění dokumentu pak vyžaduje určitou míru odvozování, což je nutně vyžadováno také od vyhledávacího stroje. Toto odvozování může produkovat nová fakta, omezení a metadata, která mohou být přínosná pro indexování. Konvenční vyhledávací stroje se o porozumění obsahu dokumentů nesnaží, protože se jedná o obtížnou úlohu a vyžaduje výzkum v oblasti porozumění textu.
4. Významně se rovněž liší struktura grafu, která je daná kolekcí dokumentů tvořících sémantický web, od grafu plynoucího z hypertextových odkazů běžných webových

dokumentů. Tento rozdíl ovlivňuje jak strategii pro automatizované získávání dokumentů, tak způsob určení metrik pro řazení jejich významnosti.

Při vyhledávání porovnává vyhledávací stroj dotazy vůči datům, která vzešla z indexování. Zatímco výsledkem u webových vyhledávacích strojů jsou dokumenty, výsledky vyhledávacího stroje sémantického webu mohou poskytovat i více než prostý dokument – lze agregovat data na více úrovních, počínaje univerzálním grafem všech RDF dat na webu až k jednotlivému RDF tripletu.



Obrázek 7: Architektura vyhledávacího stroje swoogle

Velmi podrobně se struktuře vyhledávacího stroje pro sémantický web věnuje článek [4].

5 Místo závěru – vše je otevřené

Použití označení Web 2.0 jakoby se stalo startovním výstřelem – a najednou se objevují *Business 2.0*, *Library 2.0*, *Marketing 2.0*, ... Je to vcelku pochopitelné, protože se prakticky velmi rychle ukázalo, že orientace na uživatele ve smyslu spolutvůrce a nikoliv jen „konzumenta“ obsahu je správná. Úspěch wiki, blogů, webů jako jsou flickr.com, myspace.com či youtube.com je toho nejlepším důkazem.

5.1 Web 3.0

Pojem *Web 3.0* použil poprvé počátkem roku 2006 Jeffrey Zeldman na svém blogu [19], když kritizoval Web 2.0 a jeho technologie. V těchto souvislostech se začalo hovořit o

databázové transformaci webu, o využití idejí sémantické webu, o zapojení třetí dimenze do navigace apod. Diskuse o vymezení principů Webu 3.0 se odehrává jak v rovině akademické, tak v rovině čistě komerční. Jedním ze zjednodušených a dobře srozumitelných vymezení je:

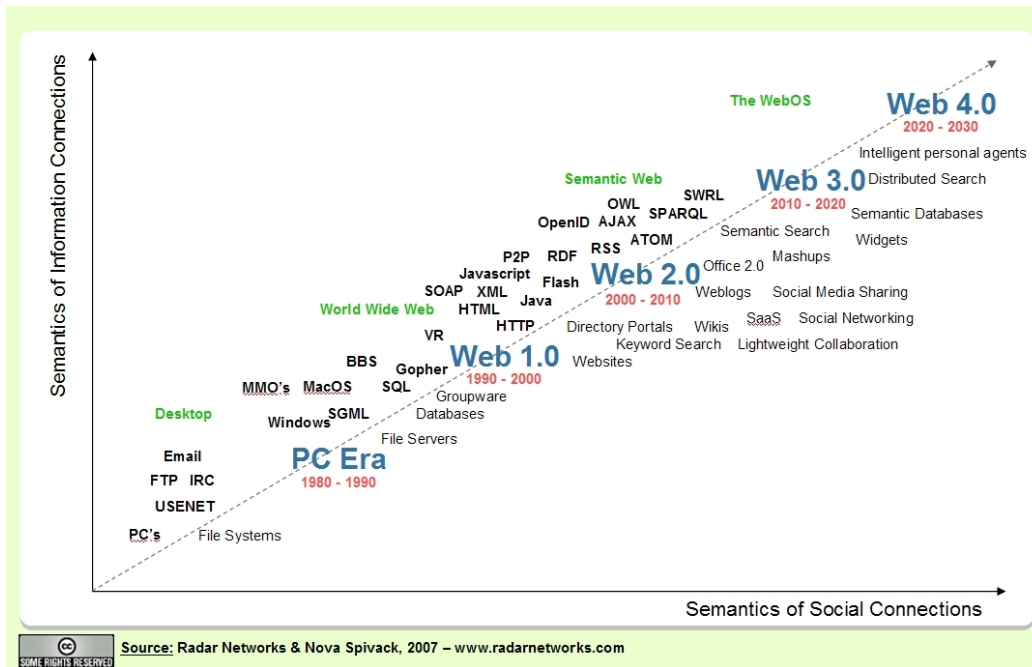
$$\text{Web 2.0} + \text{sémantický web} = \text{Web 3.0}$$

Mezi technologie, které jsou uvažovány pro použití v rámci Webu 3.0, patří:

- Artificial intelligence (umělá inteligence)
- Automated reasoning (automatické odvozování)
- Cognitive architecture (kognitivní architektury)
- Composite applications (kompozitní aplikace)
- Distributed computing (distribuované výpočty)
- Human-based genetic algorithms (genetické algoritmy)
- Knowledge representation (reprezentace znalosti)
- Ontology (ontologie ve smyslu počítačových věd)
- Scalable vector graphics (škálovatelná vektorová grafika)
- Semantic Web (sémantický web)
- Semantic Wiki (sémantické wiki)
- Software agents (softwarové agenty)

5.2 Web 4.0, ...

Ačkoliv ideje Webu 2.0 začínají přinášet první ovoce a Web 3.0 je „narýsován“ pouze v hrubých obrysech, objevují se již i vize o Webu 4.0. Protože se však jedná o předvídání, co bude zhruba za 15 let – viz obr. 8 – je velmi obtížné odlišit, co je reálné a co je spíše přání...



Obrázek 8: Rekapitulace historie vývoje na cestě k webu a směřování k Webu 4.0

Reference

- [1] Berners-Lee, Tim. SW and web 2.0 = ISWC. 2006. URL <http://www.w3.org/2006/Talks/1108-swui-tbl/>.
- [2] Berners-Lee, Tim; Hendler, James; Lasila, Ora. The semantic web. *Scientific American*, 2001, roč. 284, s. 35–43. URL <http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&catID=2>.
- [3] Clark, Kendall. SPARQL: Web 2.0 meet the semantic web. 2005. URL http://www.oreillynet.com/xml/blog/2005/09/sparql_web_20_meet_the_semanti.html.
- [4] Finnin, Tim; Ding, Li. Search engines for semantic web knowledge. In *XTECH 2006 "Building Web 2.0"*. 2006. URL <http://xtech06.usefulinc.com/schedule/paper/102>.
- [5] Geroimenko, Vladimir. *Dictionary of XML Technologies and the Semantic Web*. Springer Professional Computing. London: Springer, 2004. ISBN 1-85233-768-0.
- [6] Gruber, Tom. Where the social web meets the semantic web. 2007. URL <http://tomgruber.org/writing/social-meets-semantic-web.htm>.
- [7] Knight, Charles. Top 100 alternative search engines. 2007. URL http://www.readwriteweb.com/archives/top_100_alternative_search_engines_mar07.php.
- [8] MacManus, Richard. Web 2.0 and semantic web: Mars and venus? 2005. URL <http://blogs.zdnet.com/web2explorer/?p=13&part=rss&tag=feed&subj=zdblog>.
- [9] Metz, Cade. Web 3.0; the internet is changing... again. *PC Magazine*, 2007, roč. 26, č. Apr 10.

- [10] Nixon, Lyndon J.B. Multimedia, web 2.0 and the semantic web: A strategy for synergy. In *First International Workshop on Semantic Web Annotations for Multimedia (SWAMM)*. 2006. URL <http://image.ntua.gr/swamm2006/resources/paper14.pdf>.
- [11] OEDb. Top 25 web 2.0 search engines. 2007. URL <http://oedb.org/library/features/top-25-web20-search-engines>.
- [12] O'Reilly, Tim. What 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>.
- [13] Schraefel, M. C.; et al. Semantic web meets web 2.0 (and vice versa): The value of the mundane for the semantic web. 2006. URL <http://eprints.ecs.soton.ac.uk/12665/01/iswc06-web2point0mc.pdf>.
- [14] Sklenák, Vilém. Sémantický web. In *Inforum 2003*. Albertina icome Praha, 2003. URL http://www.inforum.cz/inforum2003/prispevky/Sklenak_Vilem.pdf.
- [15] Tenenbaum, Jay M. AI meets web 2.0: Building the web of tomorrow, today. *AI Magazine*, 2006, roč. 27, č. 4, s. 47–68.
- [16] Wikipedia. Web 2.0. 2007. URL http://en.wikipedia.org/wiki/Web_2.0.
- [17] Wikipedia. Web 3.0. 2007. URL http://en.wikipedia.org/wiki/Web_3.0.
- [18] Zambonini, Dan. Is web 2.0 killing the semantic web? 2005. URL http://www.oreillynet.com/xml/blog/2005/10/is_web_20_killing_the_semantic.html.
- [19] Zeldman, Jeffrey. Web 3.0. 2006. URL <http://www.alistapart.com/articles/web3point0>.