

PŘEDNÁŠKA 1.

Úvod do sémantického webu



Martina Husáková

13. 2. 2014

Obsah přednášky

- Vznik celosvětové sítě WWW
- Technologie využité pro WWW
- Aktuální problémy současného webu
- Možnosti řešení problémů a budoucnost

2

A řekl Tim: „Budiž WEB!“

- Myšlenka služby WWW (zkráceně Web) se zrodila v CERNu (Evropské organizaci pro jaderný výzkum), resp. v mysli vědeckého pracovníka T. B.-Leeho
- Prvotní motivace: automatické sdílení informací mezi vědeckými pracovníky z řad fyziků CERNu, kteří byli často rozmístěni po různých koutech světa
- Původní řešení: sdílení informací pomocí e-mailů a výměny souborů (ftp)

[1] Obr. URL: <http://fphe.epfl.ch/mneedham/>[2] Obr. URL: <http://www.ayushkumar.com/tim-berners-lee-father-of-html>

3

Počátky hypertextu

- Hlavní myšlenka:
„The dream behind the Web is of a common information space in which we communicate by sharing information.“ [3]
- Hlavní zdroj inspirace: koncept hypertextu prvotně zmíněný v článku V. Bushe „As We May Think“ (r. 1945)
- Návrh konceptuálního stroje Memexu

[3] Citace URL: <http://www.w3.org/People/Berners-Lee/ShortHistory.html>

4

Memex (*Memory Extended*)

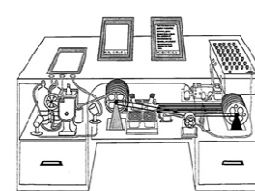
- **Memex:** systém teoreticky schopný uložit obrovské množství informací s možností tvorby relací mezi nimi samotnými uživateli
- Relace dostupné i v budoucnu pro tvorbu relací dalších (metoda tvorby asociací – velmi blízké způsobu ukládání informací v mysli člověka)
- Memex nebyl nikdy implementován, ale stal se inspirací pro T. H. Nelsona k vytvoření moderní verze hypertextu

5

Koncept hypertextu



Článek z roku 1945



Návrh stroje Memex

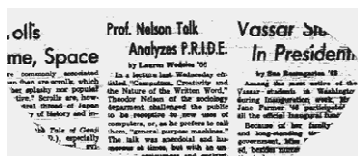
⇒ <http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/303881/>

[4] Obr. URL: <http://newmedia.wikia.com/wiki/Memex>

6

Vznik World Wide Webu (WWW)

- Pojem „**hypertext**“ poprvé zmíněn právě T. H. Nelsonem v přednášce: „Computers, Creativity, and the Nature of the Written Word“ (r. 1963)



[5] Obr. URL: http://www.livinginternet.com/w/wi_nelson.htm

7

Docuvers

- Vize zahrnovala implementaci tzv. „**Docuversu**“: systém ukládající data jen jednou bez jejich vymazávání s možností přístupu k nim pomocí linků (kniha Literary Machines)
- Inspirace pro T.B.-Leeho až s mírně odlišnými změnami resp. (odstraňování dokumentů, uložení informací i na více místech)

8

Prototyp Enquire – předchůdce WWW

- Prototyp **Enquire**: systém pro používání a sdílení dokumentů jen v rámci CERNu (r. 1980)
 - Klíčové rysy: provozuschopnost na různých typech počítačů, kompatibilita různých formátů dat, kódovacích schémata a typů počítačových sítí, použití vzájemných odkazů (hypertextu)
 - Bližší spíše wiki než aktuální podobě webu
 - Předchůdce dnešního webu
- Následná snaha o vývoj nového systému (přístupu) veřejně dostupného s možností tvorby informací nezávisle na ostatních uživateli
- Web měl za úkol sjednotit různé již existující systémy v CERNu (Enquire, Usenet, a další)

9

„HyperText and CERN“

- Specifikace návrhu vývoje webu v dokumentu „HyperText and CERN“ (r. 1989), ve kterém byla navržena architektura (struktura) webu
- Založena na myšlence tzv. univerzální čtenářské obce (universal readership):
„If information is available, then any (authorized) person should be able to access it from anywhere in the world.“ ^[6]
- Postavena na modelu klient-server

[6] Citace URL: <http://teltsar.com/web-history/>

10

Architektura WWW

- Klient**: program (zpravidla internetový prohlížeč), který se připojuje ke vzdálenému počítači (serveru) s úmyslem žádat data, informace (obecně služby) ve formě HTTP požadavku
- Server**: program ve výpočetním systému poskytující data ve formě HTTP odpovědi

11

Klíčové technologie WWW

- T. B.-Lee v návrhu („HyperText and CERN“) specifikoval klíčové technologie pro výstavbu webu, které byly ověřovány v laboratořích CERNu
 - HTML** (HyperText Markup Language)
 - HTTP** (HyperText Transfer Protocol; ver. 0.9)
 - Program pro chod HTTP serveru
 - Webový server** (první adresa a stránky <http://info.cern.ch>)
 - Webový prohlížeč** (i editor v jednom – WorldWideWeb (později Nexus))
- Plná provozuschopnost všech součástí spolu s veřejným přístupem r. 1991

12

Webový prohlížeč

- Obecně: program pro získávání, interpretaci a zobrazování dat (výsledků)
- WorldWideWeb prohlížeč kompatibilní s různými platformami => přístup k informacím z různých typů počítačů



[7]

[7] Obr. URL: <http://www.w3.org/People/Berners-Lee/WorldWideWeb.html>

13

Zajímavost

- První obrázek přístupný poprvé veřejně na webu: Les Horribles Carnettes („The Horrible CERN Girls“)



[20]

[8]

[8] Obr. URL: <http://musicclub.web.cern.ch/MusicClub/bands/carnettes/firstband.html>[20] Obr. URL: http://www.dailygalaxy.com/my_weblog/2011/12/cerns-lhc-paused-to-make-major-announcement-did-they-locate-the-higgs-boson.html

14

Zajímavost

- Webový prohlížeč WorldWideWeb běžel zprvu jen na počítačích typu NeXT s operačním systémem NeXTSTEP, obdobně i první server



← Pracovní stanice NeXT
první webový server na světě

„Tento stroj je server, nevypínat!“

[9] Obr. URL: <http://www.brighthub.com/internet/web-development/articles/123425.aspx>

15

Web 1.0

- První implementace webu
- Systém propojených hypertextových dokumentů přístupných pomocí Internetu
- **Read-only web** dle T. B.-Leeho
- Nabídka statických webových stránek
- Bez možnosti interakce mezi tvůrci informací, čtenáři nebo oběma skupinami
- Nepravidelné aktualizace informací
- Typická aplikace: osobní webové stránky
- Zaměřen na společnosti
- Cíl splněn: nabídka informací komukoliv, kdykoliv, odkudkoliv

[10] Naik, U., Shivalingaiah, D. Comparative Study of Web 1.0, Web 2.0 and Web 3.0. International CALIBER, 2008

[11] URL: <http://www.jabnol.org/internet/web-3-concepts-explained/8908/>

16

Web 2.0

- **Read-write web** (load, download)
- Zaměřen na tvorbu obsahu
- Interakce s jinými uživateli na webu
- Sdílení obsahu
- Typické aplikace (blogy, wikipedia, RSS, XML, online webové služby (Gmail, eBay), tagging („folksonomie“), Google, Flickr, YouTube)
- Zaměřen na komunity, resp. spolupráci

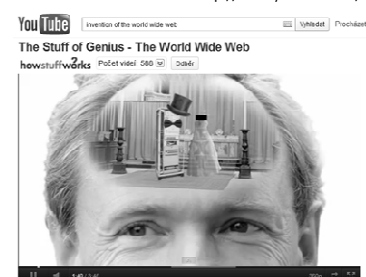
[10] Naik, U., Shivalingaiah, D. Comparative Study of Web 1.0, Web 2.0 and Web 3.0. International CALIBER, 2008

[11] URL: <http://www.jabnol.org/internet/web-3-concepts-explained/8908/>

17

Rekapitulace

Videoukázka z YouTube.com: <http://www.youtube.com/watch?v=Bkd3VyQMBK0>



18

Aktuální problémy současného webu

- Začneme jedním příběhem aneb bylo nebylo ☺ ...
 - Argentinský spisovatel J. L. Borges: Povídka Babylonská knihovna (**The Library of Babel**, 1941)
 - Knihovna obsahující všechny informace a znalosti světa, které kdy byly nebo budou napsány, přičemž neexistuje žádný návod, vodítko, jak nalézt tu námi požadovanou knihu (neexistence přístupových metod)
 - I když bychom dokument našli, není zaručeno, že mu budeme rozumět. Knihovna totiž obsahuje dané dílo vytvořené ve všech možných jazycích
 - Jak si knihovnu představit?
 - <http://www.youtube.com/watch?v=mmUGfUygyjA>
 - <http://www.youtube.com/watch?v=30Xlpuid2ME>

19

Úryvek z povídky Babylonská knihovna

- „Když se rozhlásilo, že knihovna obsahuje všechny knihy, vyvolalo to nejprve pocit podivného štěstí. Všem připadalo, jako by byli pány nedotčeného a tajného pokladu.“
- „Nehoráznou naději přirozeně vystřídala nesmírná sklíčenost. Jistota, že některý regál v některém šestiúhelníku skrývá drahocenné knihy a že ty knihy jsou nedostupné, se zdála nesnesitelná.“

[12] Citace: Sklenák, V. Vyhledávací nástroje v prostředí Internetu – co bude dál?
URL: http://platan.vc.cvut.cz/akp2003/sbornik/03_sklenak.pdf

20

Obecný problém

- K nejčastějším aktivitám uživatelů webu patří vyhledávání informací
- Vyhledání relevantní informace za přijatelný čas je jedním z nejzávažnějších problémů
- Dokument může požadované slovo či téma obsahovat, ale vyskytuje se v něm v jiném kontextu
- Je zde problém na straně aplikací, které mají problém s „pochopením“ významu webového obsahu, resp. s tzv. sémantikou

21

Detailnější pohled: Jazyk HTML (1)

- HTML (HyperText Markup Language) – příklad kódu

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="cs" lang="cs">
  <head>
    <meta http-equiv="content-type" content="text/html; charset=utf-8"/>
    <meta http-equiv="Content-Language" content="cs"/>
    <title>Nielsen</title>
    <meta name="robots" content="all, follow"/>
    <meta name="description" content="nielsen"/>
    <meta name="keywords" content="nielsen, umělecké potřeby, umělecké materiály,
      rámování"/>
  </head>
  <div id="roof">
    <h1 class="logo"><a href="/cs" title="nielsen">nielsen</a></h1>
    <ul class="tiny_menu">
      <li class="first"></li>
      <li>Špičkové rámovací materiály</li>
      <li class="space"></li>
      <li>umělecké potřeby</li> ...
    </ul>
  </div>
```

22

Detailnější pohled: Jazyk HTML (2)

- Jakou informaci vidí člověk?



[13] Obr. URL: <http://www.nielsen.cz/cs/ramovaci-materialy/aluminovy-program/alu-ramovaci-profil>

23

Detailnější pohled: Jazyk HTML (3)

- Jakou informaci „vidí“ stroj?

```
...
<div id="roof">
  <h1 class="logo"><a href="/cs"
title="nielsen">nielsen</a></h1>
  <ul class="tiny_menu">
    <li class="first"></li>
    <li>♥=♣♠♦Y </li>
    <li class="space"></li>
    <li>♥=♣♠♦Yf </li>
  </ul>
</div>
```

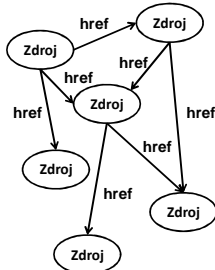
24

Problémy syntaktického webu (1)

Současný web je obvykle tvořen uzpůsoben tak, aby byl srozumitelný především pro lidi, nikoliv pro stroje.



[14] <http://www.csfd.cz/>



[15] <http://www.csfd.cz/zebricky/nejslepsi-200-filmu/>

25

Problémy syntaktického webu (2)

• V současné podobě webu je problematické:

- klást komplexní dotazy využívající znalosti
 - „Najdi zverimex v Hradci Králové, který má otevřeno denně od 10:00 do 16:00 hod. a nabízí krmivo pro psy s názvem Pro Plan Adult (1 – 10 kg) v ceně do 350,- Kč.“
- získávat znalosti z různých datových zdrojů
 - „Hledej takové dopravní spojení, abych se dostal(a) ze Smiřič do kina v Hradci Králové půl hodiny před počátkem filmu Atlas mraků.“
- delegovat komplexní úkoly webovým „agentům“
 - „Zarezervuj pobyt na teplém místě, ne příliš daleko, kde se lze domluvit anglicky.“

[16] URL: Patel-Schneider, P. F. Tutorial on the W3C OWL Web Ontology Language. ENC 2004

26

Jazyk XML jako řešení?

- XML = eXtensible Markup Language
- Značkovací jazyk definovaný konsorciem W3C roku 1998
- Účel jazyka XML:
 - pro anotaci („označování“) textových dat pomocí vlastních značek (tagů)
 - pro dodání jasné struktury webovým dokumentům
 - zachytit význam slov mezi tagy
- Dobře zpracovatelný lidmi, stroji také, ale ...

27

Detailnější pohled: Jazyk XML (1)

• Jakou informaci vidí člověk?

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<?xml-stylesheet type="text/css" href="cd_catalog.css"?>
<CATALOG>
<CD>

  <NAZEV-CD>Harem</NAZEV-CD>
  <UMELEC>Sarah Brightman</UMELEC>
  <DISTRIBUCE-CR>Panther</DISTRIBUCE-CR>
  <CENA-CD>416 Kč</CENA-CD>
  <ROK>2003</ROK>
  <POCET-SKLADEB>14</POCET-SKLADEB>

</CD>
...
</CATALOG>
```

Webová stránka



[17] http://superstarshop.cz/obchod.php?prodej=118273&rightman_sarah_harem

28

Detailnější pohled: Jazyk XML (2)

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<?xml-stylesheet type="text/css" href="cd_catalog.css"?>
<CATALOG>
  <CD>
    <NAZEV-CD> ♥=♥♥♥♥♥</NAZEV-CD>
    <UMELEC>S♥=♥♥♥♥♥</UMELEC>
    <DISTRIBUCE-CR>P ♥♥♥♥♥</DISTRIBUCE-CR>
    <CENA-CD> ♥♥♥♥♥ K</CENA-CD>
    <ROK> ♥♥♥♥♥</ROK>
    <POCET-SKLADEB> ♥♥♥♥♥</POCET-SKLADEB>
  </CD>
  ...
</CATALOG>
```

- Zachytíme sémantiku textu vlastními tagy, ale stroj stále nedovede interpretovat jejich význam

29

A řekl Tim: „Budiž sémantický web.“

- Vizi sémantického webu představil sir T. B.-Lee s kolegy v časopise Scientific American roku 2001
- Motivace: obrovské množství informací na webu, které je lehce přístupné pro člověka, ale obtížné pro „porozumění“ stroji (neexistence jednoduše přístupného formátu)

Pozn. Český překlad textu i v kurzu v Olivě!

[18] Obrázek Tim B.-Lee. URL: <http://www.radford.edu/~jennings15/text.htm>

[19] Obrázek obálky časopisu SA. URL: <http://www.scientificamerican.com/sciammag/tcontents-2001-05>

30

Iniciativa sémantického webu

- „Sémantický web by měl dodat strukturu a význam obsahu webových dokumentů, vytvořit prostředí, kde softwarové agenty přecházejí z jedné webové stránky na druhou a plní sofistikované úlohy pro uživatele.“
„Sémantický web nebude oddělen od webu současného, ale bude ho rozšiřovat tak, že informace budou mít jasný význam dovolující počítačům a lidem spolupracovat.“

T. B.-Lee, Scientific American, 2001.

[20] Citace URL: Scientific American. URL: <http://www.scientificamerican.com/sciammag/7contents-2001-05/>

31

Cíle sémantického webu

- Dělat za uživatele triviality, které by je mohly stát mnoho často drahocenného času i peněz
- Dodat informacím na webu jednoznačně reprezentovaný význam, zachycený v jazyku přístupnějším pro automatické zpracování
- Samostatně nalézat relevantní zdroje informací, zpracovávat je a odvodit ze stávajících znalostí znalosti nové
- Podpořit sdílení informací a znalostí s ostatními

[21] Kvasnička V., a kol. Umeleá inteligencia a Kognitivna veda II. Slovenská Technická Univerzita v Bratislave, ISBN 978-80-227-3284-0

32

Web 3.0

- „The portable personal web“, „Semantic Web“, „read-write execute web“
- Tvorba vysoce kvalitního obsahu a služeb s využitím technologií Webu 2.0
- Využití prvků umělé (výpočetní) inteligence, technologií 3D (3D Web) – virtuálních prostředí (s avatary)
- Nabídka různých pohledů na stejná data
- Zaměření se na kontext a personalizaci – obecněji na jedince
- Příklady aplikací: iGoogle, Google Maps, Facebook, Swoogle, FOAF

[10] Naik, U., Shivalingaiah, D. Comparative Study of Web 1.0, Web 2.0 and Web 3.0. International CALIBER, 2008

[11] URL: <http://www.latinol.org/internet/web-3-concepts-explained/8908/>

33

Web 1.0 vs. Web 2.0 vs. Web 3.0

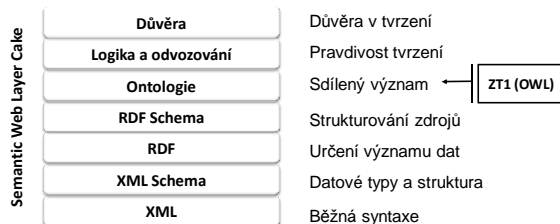
Videoukázka z YouTube.com: http://www.youtube.com/watch?v=n3sbwnFbm_s



34

Technologie sémantického webu

- Sémantický web jako soubor technologií obohacující současný web o metadata popisující význam webového obsahu



35

Příští přednáška dne 20. 2. 2014 na téma

METADATA

36