



# Základy a principy World Wide Web-u

PIA 2012/2013  
Přemysl Brada  
Ondřej Rohlík

# Co se dozvíte

- Myšlenky a technologie v pozadí webu
- Hypertext
  - klíčové vlastnosti
- Vyznačovací jazyky
  - význam, druhy
- Architektura webu
  - principy a motivace

# Základy, na nichž stojí web

- Internet \*1961
  - iniciativa US DOD: ARPA 1957 – protokoly, servery
- Hypertext \*1945, 1965
  - Vannevar S. Bush: *As We May Think* – provázání informací
  - Ted Nelson
- Logické vyznačování \*1969, 1986
  - Charles Goldfarb (IBM): SGML – strukturování textu/info
- WWW \*1990
  - Tim Berners-Lee (první návrh: CERN 1989)



# Internet: infrastruktura pro WWW

- Internet
  - » 1961 iniciativa US DOD: ARPA 1957 – protokoly, servery
  - » problém se jménem: Internet vs. internet
  - vzájemně propojené sítě => protokoly
  - technologie, operační systémy, klientské programy
  - služby, informace, lidé
- Hlavní principy, kterými se řídí
  - heterogenita, reuse, škálovatelnost
  - standardy ověřené několika nezávislými implementacemi
  - důraz na otevřené standardy

# Hypertext: Motivace

- **Problém:** *Vannevar Bush, 1945*
  - objem dostupných informací enormní
  - způsoby jak je prohledávat je primitivní
- **Lidská paměť a myšlení používá asociace**
  - psaní: **myšlenky** → **struktura** → **sekvence**
    - » sekvenčnost vynucena médiem (papír)
  - čtení: **sekvence** ⇒ **přeskakování, větvení, rejstříky**
  - síťové uspořádání dat blízké kognitivním modelům  
⇒ *Memex*
- **Trails** – propojení souvisejících konceptů
  - zapamatovat si již naučené, cestu učení

stezka, trasa

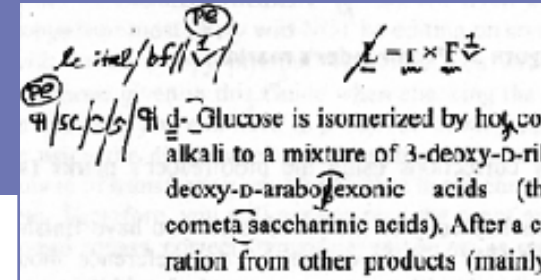
# Co je Hypertext

termín: Ted Nelson, 1965

- Klíčové principy
  - nelineární organizace textu
  - informace, koncepty: uzly
    - » nejen text ⇒ hypermedia
  - vztahy: propojení, odkazy
- Hypertextové systémy
  - softwarová podpora: GUI, aktivní odkazy, tvorba dokumentů
  - standardní mechanismy vyhledávání (klíč. slova, fulltext)
  - databáze → dokumentů, odkazů

Důsledek: klikněte [zde](#)

# Vyznačovací jazyky



- Motivace
  - publishing: 30% času stráveno formátováním
  - nové zdroje dat: CD/DVD/etc, WWW, databáze
    - 92 mld nových dokumentů / rok (USA)
  - Umíme efektivně
    - v nich hledat?
    - vyměňovat, sdílet a zpracovávat je?
- Vyznačování
  - původ: sazečské značky v rukopisu
  - elektronické dokumenty: formalizace značek, jazyky

# Fyzické vs. logické vyznačování

- Fyzické (procedurální)

*Značky definují formát zobrazení*

- ztráta informace o struktuře textu
- problémy: změna prezentace, konverze do jiného
- často technologická závislost, proprietární

- Logické (popisné, generické) – 1960+

*Značky popisují strukturu informace*

- význam částí struktury, přesnost
- výhody: umožňuje zpracování
- nezávislé na technologii, future-proof

```
\pard\plain \s1\sb240\sa60\keepn
\b\f5\fs28\lang2057\kerning28
Heading
\par \pard\plain
\f4\fs20\lang2057 LINE WITH
CAPITALS
\par {\pntext\pard\plain
\f1\fs20 \b7\tab}
```

```
<book>
<title>Getting started with
SGML</title>
<!-- hide some -->

<chapter>
<title>The Challenge</title>
<para>The explosive success...
</para>
```



# Fyzické vyznačení: RTF

```
{\rtf1\ansi \deff4\deflang1033
  {\fonttbl {\f1\froman\fcharset2\fprq2 Symbol;}
  {\f4\froman\fcharset0\fprq2 Times New Roman;}
  {\f5\fswiss\fcharset0\fprq2 Arial;} }
...
\paperw11906\paperh16838 \widowctrl\ftnbj\aenddoc\formshade \fet0\sectd
\linex0\headery709\footery709\colsx709\endnhere
...
\pard\plain \s1\sb240\sa60\keepn \b\f5\fs28\lang2057\kerning28 Heading
\par \pard\plain \f4\fs20\lang2057 LINE WITH CAPITALS
\par {\pntext\pard\plain\f1\fs20 \'b7\tab}
\pard \fi-283\li283 {\*\pn \pnlvlblt
\pnf1\pnstart1\pnindent283\pnhang{\pntxtb \'b7}} BULLET
\par {\pntext\pard\plain\fs20 1.\tab}
\pard \fi-283\li283
{\*\pn \pnlvlbody\pndec\pnstart1\pnindent283\pnhang{\pntxta .}} NUMBER
```

# Logické vyznačení: DocBook

```
<book>
<bookinfo><bookbiblio>
<title>Getting started with SGML</title>
<authorgroup><corpauthor>ArborText, Inc.</corpauthor>
<!-- hide some --></authorgroup>
<pubdate>18 October 1995</pubdate>
</bookbiblio></bookinfo>

<chapter>
<title>The Business Challenge</title>
<para>The explosive success of the Internet is an obvious
example ...
</para>
<sect1>
<title>Information Creation</title>
<para>By some estimates, 20% of our GNP is spent on generating
new information.
</sect1> <!-- hide the rest -->
</book>
```

# SGML

## *Structured Generalised Markup Language*


- Meta-jazyk pro definování logických vyznačovacích jazyků
  - nástroj pro popis gramatik ⇒ vyznač. jazyky jsou SGML aplikace
  - struktura dokumentu (elementy) vyjádřena pomocí značek
  - definovány v Document Type Definition (DTD)
  - ⇒ terminologie, produktivita, reuse, flexibilita
- Historie
  - 1960s Charles Goldfarb (IBM) první idea, 1969 GML, DTD
  - 1980 ANSI/ISO draft, 1986 ISO standard No. 8879 (ČSN 28879)
- Hlavní SGML aplikace
  - elektronické rukopisy (velká vydavatelství)
  - LinuxDoc, DocBook, HTML; XML

nikoli „tagů“  
nebo  
„příkazů“!

# XML

## *Extensible Markup Language*

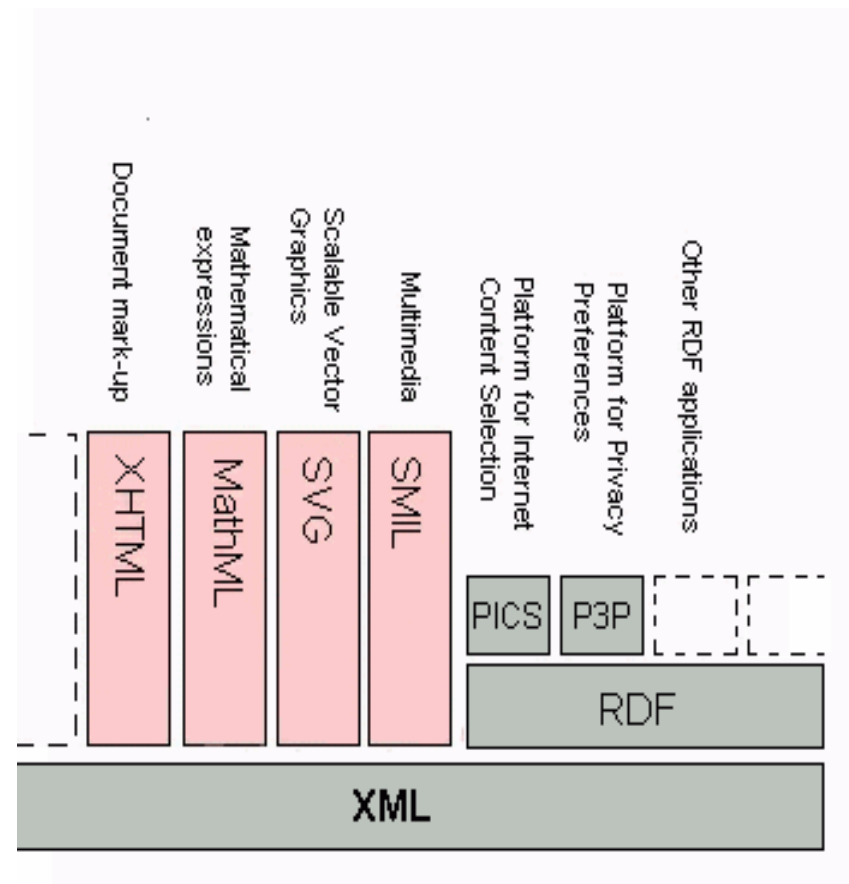
- Meta-jazyk pro definování logických vyznačovacích jazyků
  - nástroj pro popis gramatik  
⇒ vyznač. jazyky jsou XML aplikace
  - struktura dokumentu (elementy) vyjádřena pomocí značek
  - definovány v Document Type Definition (DTD)  
⇒ validace
- Terminologie, produktivita, reuse, flexibilita
- Cíle
  - generování, zpracování, transformace
  - podpora různých aplikací (text, data), Internet
  - není požadavek na stručnost zápisu



nikoli „tagů“  
nebo  
„příkazů“!

# XML standardy

- XML (2006 4th ed. & 1.1 2e)
  - 1960s -> 1980 SGML, DTD
  - 1996 Working Group at W3C
  - 1998 verze 1.0
- XML NameSpaces, InfoSet
- Doplnkové std
  - XSL, XSD
  - XQuery, XPath
- Aplikace
  - SVG, RSS, XMI, ...
  - XHTML



# World Wide Web


- “Svět informací dostupných po síti”
  - globální rozměr, umožňuje další vývoj
  - poměrně efektivní co do využití zdrojů
  - snadnost použití, přístupnost
- Client-server architektura

# Původ WWW

- 1990 projekt sdílení informací v CERNU (Tim Berners-Lee)
  - » Information Management: A Proposal – “CERN je miniaturním modelem světa, jaký bude za pár let”
  - Problém: dohledání informací v organizaci, jejíž struktura se průběžně proměňuje (⇒ pevná informační struktura neadekvátní)
  - Řešení: **de-centralizovaná**, heterogenní databáze dokumentů, **hypertextové** odkazy (síťová struktura info), zpřístupnění stávajících dat
    - » Re-use of zajímavých myšlenek: internet, hypertext, vyzn. jazyky



# Designové charakteristiky

- Klíčové prvky designu WWW
  - URI adresy, (jednosměrné) odkazy
  - objekty s různým typem obsahu
- Kvalitativní aspekty 
  - jednoduchost, modularita, tolerance
  - decentralizace, jednoduchá integrace (TOII)
  - least power, otevřenost vývoji (princip částečné srozumitelnosti)
  - plus vše co je zděděno z vlastností Internetu



# Sít' technologií

- Základní
  - spolehlivý přenos dat – TCP
  - dodací protokoly – HTTP, BASE64
  - obsahové formáty – HTML, XML, +any
  - indikace formátu – MIME
  - zabezpečení – SSL, TLS
- Přidané
  - multimedia – obrazové formáty, VRML, Flash/SMIL, video
  - vzdálené aplikace/služby – XML, SOAP
  - skriptování na klientu – JavaScript
  - provázání dalších technologií – SMS, VoIP

# Architektura Webu: URI ... URL, MIME

- URI = Uniform Resource Identifier

- klíčový prvek vyjadřující architekturu webu
- celosvětově unikátní adresa

- » W3C doporučení česky: <http://www.kosek.cz/w3c/webarch/#pr-global-id>
- » [RFC 3986](#) ... např. sekce 3

- struktura

- » <http://www.server.cz/podprostor/object.ext?param=val&next=%0A>
- » <http://ajaxian.com/archives/uri-vs-url-whats-the-difference>
  - příklad: [http://www.w3.org/Icons/w3c\\_home](http://www.w3.org/Icons/w3c_home)

URI

<http://weather.example.com/oaxaca>

Identifies

Resource

Oaxaca Weather Report

Represents

Representation

```
Metadata:
Content-type:
application/xhtml+xml

Data:
<!DOCTYPE html PUBLIC "...
"http://www.w3.org/...
<html xmlns="http://www...
<head>
<title>5 Day Forecaste for
Oaxaca</title>
...
</html>
```

- URL = ... Locator

- Speciální případ URI

- » umístění konkrétní reprezentace informace na síti
- » [RFC 1738](#) ... 3.3 HTTP

- *schema* označuje přístupový mechanismus

- MIME = Multipurpose Internet Mail Extensions

- std označení formátu obsahu, struktura je v RFC 2045
  - » text/html    application/xhtml+xml    text/plain; charset=UTF-8
  - » image/png , application/pdf , application/msword , x-text/textile

# Aktéři ve světě WWW

- poskytovatel informace (zdroje)
  - text, data, multimedia
  - business case, investice
- designer, vývojář (technologie)
  - informační architektura, prezentace obsahu, navigace
  - programování a skriptování
- webmaster (údržba)
  - celková zodpovědnost za website
  - skriptování, administrace, konzultace
- poskytovatel služeb (hosting, infrastruktura)
  - diskový prostor, internet, databáze
  - servery, cachování a proxy, síť, protokoly
- návštěvník, jeho **klientský software** (čtení, interakce)
  - zobrazení HTML+CSS, plug-ins



# Standardizační orgány

- W3C – World Wide Web Consortium
- Microsoft



*Leading the Web to Its Full Potential...*

- WhatWG
- OASIS (oasis-open.org)
- Adobe
- ECMA, ISO

... you!