- Cvičení jednohodinová, účast povinná. Na cvičení se registrujte!
- Kredity: tři plus ukončení.





"You cannot *not* communicate"[1]
*Paul Watzlawick (1921– )*

# Lidé komunikují prostřednictvím dokumentů připravovaných elektronicky

Petr Sojka: Materiály k výuce předmětu Elektronická příprava dokumentů (PB029) na FI MU.

*Verze z 14. 1. 2009 formátovaná k úspornému tisku.*

Petr Sojka: Digitální typografie a sázecí systém TEX
1. přednáška PB029, 15. 9. 2008, Brno © Petr Sojka, 2000–2008

## Obsah dnešní přednášky



- Základní pojmy.
- Cyklus přípravy dokumentů.
- Sylabus kursu.
- Doporučená literatura.
- Test – korektura.
- Anketa.
- Logické a vizuální značkování.
- Dotazy, agenda (cvičení, rozdělení do skupin), brožurka *Začínáme s TEXem na FI*.

Petr Sojka: Digitální typografie a sázecí systém TEX
1. přednáška PB029, 15. 9. 2008, Brno © Petr Sojka, 2000–2008

## PB029



Elektronická příprava dokumentů
Electronic Preparation of Documents

- Petr Sojka[1], mailto://sojka@fi.muni.cz
- Materiály k předmětu budou průběžně zveřejňovány na webové stránce předmětu[2], případně v IS MU.
- Aktuální sylabus předmětu[3] je na IS MUNI[4].
- Doporučená komunikace: diskuzní fórum předmětu na ISu nebo news:cz.muni.courses.PB029.
- Konzultační hodiny na podzim 2008:
  úterý 10:30–12:00 (v C523, 5. patro C, Botanická 68a)
  středa 14:00–15:50 (v B010, pro BP, DP, projekty),
  po domluvě emailem i jindy.

[1] http://www.fi.muni.cz/usr/sojka/
[2] http://www.fi.muni.cz/~sojka/PB029/
[3] https://is.muni.cz/auth/predmety/uplny_vypis?fakulta=1433;obdobi=3082;kod=PB029
[4] http://is.muni.cz

Petr Sojka: Digitální typografie a sázecí systém TEX
1. přednáška PB029, 15. 9. 2008, Brno © Petr Sojka, 2000–2008

# Základní pojmy

- ☞ Informatika – „Nauka o vzniku, oběhu a působení *informací*.“
- ☞ Informace, její předávání – **dokument**: “A **document** is one or more official pieces of paper with writing on them.” (Collins Cobuild)
- ☞ **Příprava** dokumentů: obsah a *forma* dokumentů
- ☞ Prezentace informací a její důležitost, stále častěji **elektronicky**: “An **electronic** process or activity involves the use of electronic devices.”
- ☞ Stále vyšší požadavky na **přípravu** dokumentů: “**Preparation** is the process of getting something ready for use or for a particular purpose.”





# Cyklus ladění dokumentu

“A man who would letterspace lower case would steal sheep.”
*Frederick Goudy, (1894–1945)*

“If this wisdom needs updating, it is chiefly to add that *A woman who would . . .* ”
*Robert Bringhurst, (1946–)*

- ☞ Autor ⟶ čtenář
- ☞ Cyklus ladění dokumentu: paralela s laděním programu (waterfall diagram)





① Pořízení textu, značkování logických částí
② Design – návrh dokumentu
③ Sazba, typografie
④ Korektury
⑤ Vytvoření předlohy/matrice
⑥ Tisk (vystavení) dokumentu





- ☞ Dynamické a odvozené „dokumenty“ ⟶ hypertext, interaktivní texty, e-learning, virtuální realita, . . .

„Dokumenty v TEXu jsou mstou akademiků studentům, takže přežijí do té doby, dokud budou přežívat zatvrzelci v akademických sborech (řádově několik tisíciletí).“
*Jiří Hlavenka, (1964– )*

:-( Naproti tomu:
„Rozhodně, prosím, opusťte vody nepoučených ‚nalévačů textu‘ či ‚myšařů‘ a začněte vyrábět sazbu, jíž je nejen radost číst, ale na níž i oko se zalíbením spočinouti může.“
*Michal Ptáček*

“TEX in Schools? Just Say No!” (Konrad Neuwirth)



## Sylabus přednášky



Předmětem výuky tohoto předmětu je výklad základních principů, algoritmů a technik tvorby dokumentů, s ohledem na jejich využití při publikační činnosti (příprava skript, prezentačních materiálů, psaní diplomové, dizertační práce, dokumentace programu ap.), se zaměřením na systém TEX.

① Postup při přípravě dokumentů; základní pojmy a terminologie.

② Logická struktura dokumentu; značkování, značkovací jazyky, SGML, HTML, XHTML, XML, WML. Gramatiky dokumentů, DTD, validace. NSGMLS.

③ Návrh. Principy knižního designu. Návrh strukturovaných dokumentů.

④ Publikace na síti Internet (WWW) a specifika návrhu a vytváření těchto dokumentů. CSS, XSL.

⑤ Základy typografie, specifika sazby českých textů.



## Shrnutí – Dokumenty a VT



☞ Použití výpočetní techniky při celém cyklu práce s dokumenty (vytváření, testování a údržba dokumentů – nosičů informací).

☞ Dokumenty tištěné; zaměření na technické, strukturované (DP, odborné publikace, články, přípravy).

☞ Hypertext: „dokumenty“ elektronické, jejich značkování, vytváření, validace, konverze, šíření.

☞ Přednášky – „teorie“, důraz na principy; cvičení – „praxe“, sdílení zkušeností se softwarovými systémy a el. vytvářením dokumentů, *TEX*.

☞ Hodnocení – bodovací systém: závěrečný test sestávající se z korektury (10 %) a testu (40 %). Vytvoření dokumentů tištěného (strukturovaného, alespoň 4 strany, např. životopis) a elektronického (WWW či WAP domovská stránka), domácí úlohy a dokumenty (50 %). Prémiové body.





⑥ Písma, typy písem, způsoby reprezentace a designu písem. Rastrovací algoritmy, techniky redukcí tvaru písem. Formáty písem.

⑦ Sázecí systémy. TEX jako příklad dávkového sázecího systému a jeho principy. Sazba. LATEX. Styly sazby a jejich implementace (makroprogramování).

⑧ Algoritmy řádkového a stránkového zlomu, hz-algoritmus.

⑨ Jazyky pro popis stránek. PostScript. Bézierovy křivky.

⑩ Výstupní zařízení, charakteristiky výstupních zařízení. Osvit, tisk a vazba.

❶ Hypertext, hypertextové systémy. Portable Document Format, technologie Adobe Acrobat, pdfTEX.

❷ Paralelní publikace na papíře a na síti. Publikace databází. Konverze mezi různými formáty. LATEX2html. Dynamické dokumenty.

❸ Závěrečné shrnutí, sdílení zkušeností. Vyhodnocení příkladů. Test.

## Literatura




① Donald Knuth: Digital Typography, Addison-Wesley, 1999.

② Vladimír Beran a kol.: Aktualizovaný *Typografický manuál*, Kafka, 2000.

③ Jean-Luc Dusong, Fabienne Siegwartová: *Typografie od olova k počítačům*, Svojtka, Vašut, 1997.

④ Petr Olšák: *Typografický systém TEX*, Konvoj&CSTUG, 2000.

⑤ Jiří Kosek: *XML pro každého*, Grada, 2000.

⑥ Pavel Satrapa: *WWW pro čtenáře, autory a misionáře*.

⑦ Petr Olšák: *TEXbook naruby*, Konvoj, 1998.

⑧ P. Taylor, J. Zlatuška: *Book design*, sborník SOFSEM 1993, Zpravodaj CSTUG 1–4/95.

⑨ P. Pop a kol.: Sazba I, SPN 1984.




## Relevantní kurzy na FI



- ☞ PA178 – Digital Typography and Visualization
- ☞ PV212 – Seminar on Dig. Typography
- ☞ PV126 Typographic programming (Hans Hagen, blokově listopad)
- ☞ PV123 – Základy vizuální komunikace
- ☞ PB138 – Moderní značkovací jazyky
- ☞ PV066, PV067, PV099 – Typografie I, II, III
- ☞ PV078, PV083, PV100 – Grafický design I, II, III
- ☞ PV084, PV085, PV101 – Písmo I, II, III
- ☞ PB009 – Základy počítačové grafiky
- ☞ PV113 – Softwarové elektronické publikace




⑩ Jiří Rybička: LATEX pro začátečníky, Konvoj 2003, 3. vydání.

❶ Zdenka Dvořáková: DTP a předtisková příprava, CPress 2008.

❷ Časopisy *Typografia*, *Font*, *Electronic Publishing*, *PC World, ročník 94*.

❸ Studijní materiály v ISu, odpovědníky.

❹ Odkazy na WWW stránce předmětu a dokumentech předmětu (slidy), na www.kosek.cz[6], www.CSTUG.cz[7].

❺ http://www.fi.muni.cz/~sojka/PB029/typobib.pdf[8].

❻ knihovna.muni.cz[9].


[6] http://www.kosek.cz
[7] http://www.cstug.cz
[8] http://www.fi.muni.cz/~sojka/PB029/typobib.pdf
[9] http://knihovna.muni.cz





## Relevantní kurzy na FI



- ☞ nebo i PV030 – Textové informační systémy
- ☞ příbuzné kurzy (počítačová lingvistika, multimédia, Java)
- ☞ bohatá knihovna (`ssh knihovna.muni.cz`), polička předmětu v knihovně s materiály
- ☞ brožurka Začínáme s TEXem na `fi.muni.cz`[5]
- ☞ DVD TEXlive 2008, modul texlive-2008 (2003, -2005a, -6), stránky CSTUG

[5] http://www.fi.muni.cz/~sojka/PB029/zaciname.html

## Osnova dnešní přednášky



- ☞ Anketa a agenda.
- ☞ Logické a vizuální značkování.
- ☞ Značkovací jazyky. Formální dokumenty a principy značkování – SGML.
- ☞ Gramatiky dokumentů, DTD.
- ☞ Validace a vytváření SGML dokumentů.



## Minitest – korektura



① Najděte během deseti minut samostatně v přiloženém textu deset *typografických* chyb.

② Spočtěte si počet úspěšně nalezených.



## Agenda



- ☞ Zhodnocení ankety.
- ☞ Domácí úlohy – logické/vizuální značkování: noste do cvičení.
- ☞ Registrace, zápisy do cvičení, odhlašování ap.
- ☞ Domácí a prémiové úlohy sledujte pravidelně na webu předmětu!
- ☞ Body a prezence na cvičení je ve vystavených záznamnících na IS (kontrolujte si).
- ☞ Bloková přednáška PV126 Typographic Programming! Zapisujte se! Sdílení záp. dokumentů.



## Anketa



① Dosavadní praxe vytváření *tištěných* dokumentů (zejména technických, i pracovních):

(a) rozsah a typ dokumentů (odb. lit., eseje, dopisy, . . . )
(b) jaký software používáte?
(c) v čem jste s dosavadním stavem nespokojeni?

② Totéž pro *elektronické/multimediální* dokumenty (HTML, PDF, případně alespoň elektronická pošta apod.)

③ Proč jste si předmět zapsali (co si slibujete od jeho absolvování) – *kromě kreditů*?

④ Návrhy, připomínky k zaměření, obsahu apod.?

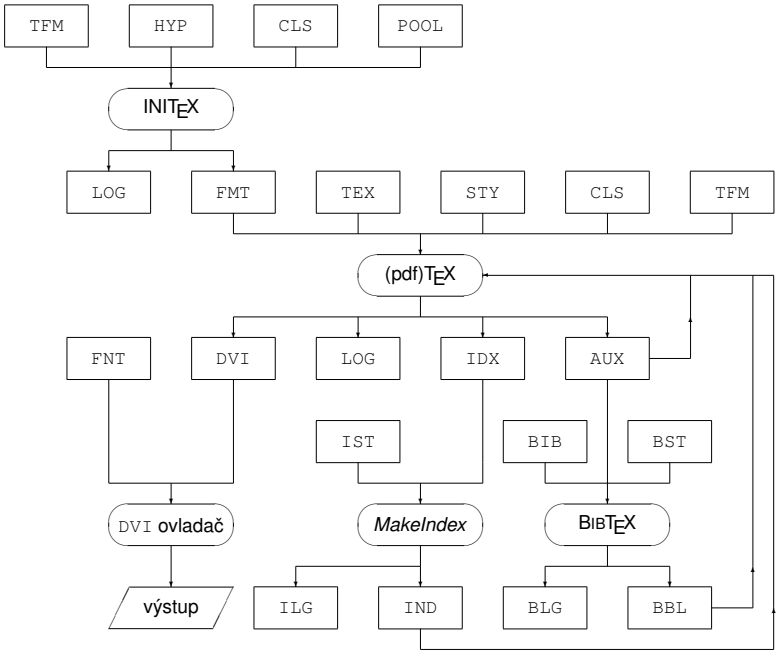
⑤ Jak hodláte zúročit informace z předmětu v praxi?

⑥ Jakou vlastníte relevantní literaturu a jaké příbuzné předměty jste absolvovali respektive máte zapsány?

⑦ Oblast expertízy, zkušeností s publikováním, používané softwarové systémy (o čem byste chtěli referovat)?

## Cyklus vývoje dokumentu v LaTeXu








## Vizuální a logická struktura dokumentu



"Algorithms + Data = Program"
*Niklaus Wirth*
„Forma + Obsah = Dokument"
*Petr Sojka*

- ☞ Přenos informace – raison d'être dokumentu
- ☞ Obsah – stylistika, jazyková správnost, koheze
- ☞ Forma
  - ❑ jednotnost
  - ❑ přenos informace
  - ❑ struktura

$$\frac{\text{typografie}}{\text{literatura}} = \frac{\text{interpretace}}{\text{kompozice}} \text{skladby}$$

- ☞ Obsah $\longrightarrow$ Forma; vizuální prvky musí podporovat vnitřní obsah a strukturu a být s ní konzistentní.
- ☞ Značkovací jazyky (**M**arkup **L**anguages): SGML, HTML, WML, XML, LaTeX umožňují oddělit obsah a formu tam, kde je to možné.



## Použití přípon souborů



| SOUBOR | POPIS | IDENTIFIKACE (SUFFIX, EXTENZE, ETC.) |
|---|---|---|
| TEX | textový vstup | tex, ltx |
| DVI | výstup (pdf)TEXu, zformátovaný text | dvi |
| LOG | log soubor (pdf)TEXu | log, lis, list |
| HYP | vzory dělení slov | tex |
| TFM | metriky fontů | tfm |
| POOL | chybová hlášení | pool, poo, pol |
| FMT | format file | fmt |
| CLS | TEXové styly | tex, cls, dtx |
| STY | TEXové pom. styly file | sty, tex, doc |
| AUX | pomocné soubory | aux, toc, lot, lof, glo, tmp, tex |
| IDX | neutříděný rejstřík | idx |
| IST | styl sazby rejstříku | ist |
| ILG | *MakeIndex* log soubor | ilg |
| IND | utříděný rejstřík | ind |
| BIB | bibliografické záznamy | bib |
| BBL | seznamy citací | bbl |
| BLG | BIBTEX log soubor | blg |
| BST | BIBTEX style soubor | bst |



- ☞ Značky logické struktury $\times$ vizuální.



- ☞ Někdy oddělit nelze (Trychtýř[10] Christiana Morgensterna).

[10] http://www.alb-neckar-schwarzwald.de/morgenstern/morgenstern_poems.html

## Jaký značkovací jazyk?



- ☞ Nevýhody *proprietárních* formátů (Word, WordPerfect): účelově se mění, jsou nestabilní. Hrozí babylonské *zmatení* jazyků nebo *monopol* nevhodného formátu.
- ☞ Organizace **ISO**[11] (International Standards Organization), vydává známé normy jako ISO 8859-2, 10646-1 (Unicode), . . .
- ☞ Norma ISO 8879:1986 Information processing – Text and office systems – **Standard Generalized Markup Language** (SGML), vydaná 15. 10. 1986 po dlouhém procesu standardizace na základě návrhu Charlese Goldfarba a jeho GML. Počátky již na konci šedesátých let při návrhu informačního systému právnických textů v IBM.
- ☞ Celá sada standardů ISO: kromě SGML, DSSSL (Document Style Semantics and Specification Language, ISO/IEC DIS 10179.2:1994), SPDL (Standard Page Description Language, ISO/IEC DIS 10180:1991) a HyTime (Hypermedia/Time-based Structuring Language, ISO/IEC 10744:1992). O

[11] http://www.iso.org/



## Z hlavy autora do elektronické podoby



"Data cannot be used at a finer grain
than it is marked up at."
*R. Jelliffe*

- ☞ Autor, jeho interní model problematiky v hlavě a jeho (neustálá) reorganizace. Lineární zápis v časovém okamžiku formou **textu** (psaní). O
- ☞ Tentýž text může mít více interpretací.
- ☞ Pro uchopení obsahu a automatizaci zpracování je nutný **značkovací jazyk**: text je obohacen a *zjednoznačněn* **značkami**.
- ☞ Příklady značek: `:-)` (ze slovníčku smileys) či `<vtip>...</vtip>`.
- ☞ **Značka** je kód přidávaný k (elektronicky) vytvářenému textu, který definuje strukturu textu (**logická značka**) nebo formát textu (**vizuální značka**).



- ☞ S odstupem času: nesporný úspěch SGML, základ pro další značkovací jazyky (XML), ale například neúspěch SPDL oproti proprietárnímu PostScriptu.



Petr Sojka: Digitální typografie a sázecí systém TEX
2. přednáška PB029, 22. 9. 2008, Brno © Petr Sojka, 2000–2008

- ☞ Značka explicitně určuje interpretaci (víceznačného) textu. Dříve sazeč určoval interpretaci implicitně z kontextu a sémantiky textu (holý text v přirozeném jazyce ve strojopise je víceznačný).
- ☞ **Procedurální** (jména pro zpracování) versus **deskriptivní značkování** (jména pro kategorizaci).
- ☞ Výhody deskriptivního značkování:
  - ❑ Nezávislé zpracování označených dat různými způsoby a programy. Tedy například umožňuje generování různých výstupů z jednoho zdroje (databáze, dobře označkovaný text) pro elektronickou (XML, HTML, PDF, Hypercard, . . . ) nebo tištěnou verzi (PS, PDF z LaTeXu).
  - ❑ Oddělení obsahu a formy (ohledně formy má rozhodující slovo nakladatel).
  - ❑ Je snazší výměna a komunikace obsahu: komunikace mezi (spolu)autory, redakcí (přes Internet).
  - ❑ Datová nezávislost: textové soubory, dlouhodobá archivace trvalých hodnot [papír (500 let) versus bity (k nové verzi programu)].

→ *lingua franca* značkovacích jazyků?

## SGML deklarace

- „Pokyny pro parser“ s definicemi abeced (analogie popisu BNF):
  - role oddělovačů;
  - rezervovaná jména;
  - zakázané znaky, typicky kontrolní ASCII;
  - pravidla pro pojmenovávání, citlivost na malá a velká písmena;
  - velikosti délek, značek, . . . ;
  - komentáře;
  - parametry chování, které vlastnosti kontrolovat (OMITTAG).
- Pokud deklarace v dokumentu není uvedena, používá se implicitně **referenční syntaxe** Reference Concrete Syntax (RCS).
- příklad deklarace[13]:

[13] ftp://ftp.ifi.uio.no/pub/SGML/declaration





## SGML – historie a motivace

- Výhody: znovupoužívání částí dokumentů, kvalita a otevřenost systému, obecnost, nezávislost na konkrétní formě (WWW konsorcium), rozšiřitelnost dle technologických možností (hypertext), snížení nákladů, možnost validace (ověření korektnosti dokumentu na základě formální definice jazyka dokumentu).
- Nevýhody: za obecnost se platí složitostí, i po šestnácti letech jsou nejlepší systémy na plné SGML drahé, změna s rozšířením HTML, přesto přechod k XML (složitost).
- Uživatelé SGML (zpočátku velké instituce, elektronický a letecký průmysl, nakladatelé): AAP, OUP, Boeing, Lufthansa, British Patent Office, Association of German editors, TEI, McGraw-Hill, LDC, . . . viz stránky Organization for the Advancement of Structured Information Standards[12] .
- Dnes přesnější a relevantnější vyhledávání na Internetu.

[12] http://www.oasis-open.org/cover/





```
<!SGML "ISO 8879:1986"
  CHARSET
  BASESET "ISO 646:1991//CHARSET
                   IRV//ESC 2/8 4/2"
  DESCSET
      0   9 UNUSED
      9   2   9         -- TAB, LF --
     11   2 UNUSED
     13   1  13         -- CR --
     14  18 UNUSED

     32  95  32
    127   1 UNUSED
  CAPACITY SGMLREF
    TOTALCAP    35000
...
 NAMECASE
        GENERAL YES
        ENTITY  NO
    DELIM
      GENERAL SGMLREF
      MDO       "<!"      -- markup decl open --
      MDC       ">"       -- markup decl close --
      DSO       "["       -- declaration subset open --
      DSC       "]"       -- declaration subset close --
      MSC       "]]"      -- marked section close --
      COM       "--"      -- comment --
      RNI       "#"       -- reserved name indicator --
      LIT       "&#34;"   -- literal --
      LITA      "'"       -- alternative literal --
      GRPO      "("       -- group open --
```





## Metajazyk SGML – definice

- SGML je mezinárodní standard pro popis značkovacích jazyků.
- Formálněji je to **syntaktický *meta*jazyk** pro definici tříd dokumentů (deskriptivních značkovacích jazyků), nezávislých na abecedě.
- Formální definice typu dokumentu se nazývá **DTD** – Document Type Definition, **definice typu dokumentu** (gramatika třídy dokumentů). SGML parser (validátor) pak čte na vstupu DTD a kontroluje formální správnost.
- Instance SGML dokumentu (dále jen SGML dokument) sestává z **deklarace** (pokynů pro parser), **gramatiky** (DTD) a vlastního popisně *označkovaného textu*.

## Deklarace elementů a atributů

- Sekvence, alternativa, `ANY`, `EMPTY`, `#PCDATA`.
- Opakování: právě jednou, nejvýše jednou (?), alespoň jednou (+), libovolněkrát (*).
- Typy atributů `CDATA`, `NMTOKEN`, `NMTOKENS`, `ID`, `IDREF`, `IDREFS`, `ENTITY`, `ENTITIES`, `výčet`.
- Implicitní hodnoty atributů: `"hodnota"`, `#REQUIRED`, `#IMPLIED`, `#FIXED "hodnota"`.





```
      GRPC      ")"       -- group close --
      AND       "&"       -- and connector --
      OR        "|"       -- or connector --
      SEQ       ","       -- seq connector --
      OPT       "?"       -- opt occurrence indicator --
      REP       "*"       -- rep occurrence indicator --
      PLUS      "+"       -- plus occ ind, inclusion --
      MINUS     "-"       -- exclusion, omission flag --
      CRO       "&#"      -- character reference open --
      ERO       "&"       -- entity reference open --
      PERO      "%"       -- parameter entity reference open --
      REFC      ";"       -- reference close --
      PIO       "<?"      -- processing instruction open --
      PIC       ">"       -- processing instruction close --
      STAGO     "<"       -- start tag open --
      ETAGO     "</"      -- end tag open --
      TAGC      ">"       -- tag close --
      NET       "/"       -- null end-tag --
...
```





## Entity

- Vhodné pojmenovat části a fragmenty SGML, některé řetězce (`&TeX;` v IS) a znaky použité jako oddělovače v gramatice: `<`, `&`, `"`.
- **Entity** interní textové, externí textové, externí binární a parametrické.
- Příklad množiny interních textových entit: ISO Latin2[14].

```
<!-- Character entity set. Typical invocation:
  <!ENTITY % ISOlat2 PUBLIC
   "ISO 8879:1986//ENTITIES Added Latin 2//EN"> %ISOlat2;
-->
<!ENTITY abreve SDATA "[abreve]"--=small a, breve-->
<!ENTITY Abreve SDATA "[Abreve]"--=capital A, breve-->
<!ENTITY amacr  SDATA "[amacr ]"--=small a, macron-->
<!ENTITY Amacr  SDATA "[Amacr ]"--=capital A, macron-->
...
```

- Externí textové entity:

```
<!ENTITY název SYSTEM "URI">
```

- Externí binární entity:

[14] ftp://ftp.ifi.uio.no/pub/SGML/ENTITIES/ISOlat2





## SGML DTD: atributová gramatika

- Terminologie teorie formálních jazyků: neterminály (**elementy**), terminály (**entity**).
- Elementy vytvářejí stromovou strukturu, nemohou se navzájem křížit, jeden element je kořenový. Lze použít prázdné elementy. `<br/>`
- Elementy mohou mít **atributy**. O

```
<!ELEMENT faktura   (odberatel,dodavatel, polozka+)>
<!ELEMENT odberatel (nazev,adresa,ico,dic)>
<!ELEMENT dodavatel (nazev,adresa,ico,dic)>
<!ELEMENT polozka   (popis?,cena,dph,ks?)>
<!ELEMENT nazev     (#PCDATA)>
<!ELEMENT adresa    (#PCDATA)>
<!ELEMENT ico       (#PCDATA)>
<!ELEMENT dic       (#PCDATA)>
<!ELEMENT popis     (#PCDATA)>
<!ELEMENT cena      (#PCDATA)>
<!ELEMENT dph       (#PCDATA)>
<!ELEMENT ks        (#PCDATA)>
<!ATTLIST faktura
          cislo     CDATA #REQUIRED
          vystaveni CDATA #REQUIRED
          splatnost CDATA #REQUIRED
          vystavil  CDATA #IMPLIED>
<!ATTLIST cena
          mena      CDATA "CZK">
```

## Připojení DTD k instanci dokumentu



- Odkazem na soubor:

```
<!DOCTYPE faktura SYSTEM "faktura.dtd">
<faktura> ...</faktura>
```

- DTD součástí dokumentu:

```
<!DOCTYPE faktura [
 <!ELEMENT faktura (odberatel,
            dodavatel, polozka+)>
 ...
 ]>
 <faktura> ... </faktura>
```

- **Veřejný identifikátor** `PUBLIC` s URL:

```
<!DOCTYPE wml PUBLIC "-//WAPFORUM//DTD
                            WML 1.3//EN"
"http://www.wapforum.org/DTD/wml13.xml">
```

V případě veřejného identifikátoru je mapování od řetězce k souboru určeno tzv. **katalogem** (proměnná okolí `SGML_CATALOG_FILES` či `XML_CATALOG_FILES`).





`<!ENTITY název SYSTEM "URI" NDATA "notace">`

Petr Sojka: Digitální typografie a sázecí systém TEX
2. přednáška PB029, 22. 9. 2008, Brno © Petr Sojka, 2000–2008

## Příklad SGML dokumentu



```
<!DOCTYPE faktura SYSTEM "faktura.dtd">
<faktura vystaveni="6.10.2000"
        splatnost="20.10.2000">
  <odberatel>
    <nazev>Ferda Pistorius</nazev>
    <adresa>Boubín 77</adresa>
    <ico>2862667777</ico> <dic>291-2862667777</dic>
  </odberatel>
  <dodavatel>
    <nazev>Hynek Bedna</nazev>
    <adresa>Máchovo jezero 78</adresa>
    <ico>2862467777</ico> <dic>291-2862467777</dic>
  </odberatel>
  <polozka>
    <popis>XML editor</popis>
    <cena mena="Kč">500</cena>
  </polozka>
</faktura>
```

Petr Sojka: Digitální typografie a sázecí systém TEX
2. přednáška PB029, 22. 9. 2008, Brno © Petr Sojka, 2000–2008

## Další příklady DTD



- DTD diplomové práce [O].
- DTD informací o studijním předmětu [O].
- DTD rozvrhu.

Petr Sojka: Digitální typografie a sázecí systém TEX
2. přednáška PB029, 22. 9. 2008, Brno © Petr Sojka, 2000–2008

## Osnova dnešní přednášky



- ☞ SGML – validace a parsing
- ☞ Výměna dokumentů na Internetu, URI/URL/URN.
- ☞ Aplikace SGML: HTML, XHTML.
- ☞ XML? XML!
- ☞ WML.



## Pohledy na SGML



- ☞ **Lexikální pohled**: SGML dokument je řetěz znaků, z nichž některé jsou **data** a některé **oddělovače**, *značky*.

```
<article>
<title>Úvod do SGML</title>
<section>SGML: hlavní principy</section>
<subsection>Zobecněné značkování</subsection>
<p>Základem ...
...
</article>
```

- ☞ **Syntaktický pohled**: SGML dokument má tři části: SGML deklaraci (instrukce pro SGML parser), DTD (gramatiku jazyka) a vlastní instanci dokumentu (v tomto jazyce). Instance dokumentu obsahuje *data* a *značky* vyhovující DTD. Tento pohled vytváří parser z lexikálního pohledu.
- ☞ **Pohled hierarchie objektů**: SGML dokument je uspořádaná hierarchie (obvykle stromová struktura) objektů s obsahem (**elementů**). Elementy mají atributy, obsah a další vlastnosti. O



## SGML – validace a parsing



- ☞ Validace. Parsery. SP a sgmls/nsgmls Jamese Clarka[15].
- ☞ Výstup parseru: **Element Structure Information Set** (ESIS) formát.
- ☞ Ukázky validace, chyb, ladění, ESIS.
- ☞ Třídy dokumentů. CATALOG. `/packages/share/sgml-catalogs/`
- ☞ Další příklady DTD – informace o studijním předmětu. O

[15] http://www.jclark.com



- ☞ **Entitně-strukturní pohled**: SGML dokument je vnořená kolekce **entit**. Většina entit je pojmenovaných. **Textové** entity mohou odkazovat na další entity, zatímco **datové** entity obsahují libovolná data.



- ☞ **Aplikační pohled**: SGML dokument je cokoli, co autor chce aby jím bylo. *Data belongs to whoever creates it, and you get to decide what's important about your own data.*

## Vytváření značkovaných dokumentů



- Běžný ascii editor (`emacs`, `joe` nebo i `notepad`).
- PSGML mód emacsu: příklad stránky předmětu.
- Komerční systémy: Arbortext (Epic), SoftQuad, Corel XMetal . . .
- (X)HTML svět: Mozilla/Firefox, MSIE, Amaya, HomeSite, Netscape, FrontPage, . . .



## Značkovací jazyky na Internetu – vývoj



- 3/1989: návrh projektu World Wide Web, Tim Berners-Lee, CERN
- 12/1990: návrh HTML DTD, první Web software pod NExT
- 1991: první WWW prohlížeč pro omezené užití
- 1992: CERN začíná propagovat WWW projekt
- 2/1993: NCSA zveřejňuje alfa verzi prohlížeče Mosaic/X Marca Andreesena
- 7/1993: HTML 1.0 specifikace (Hypertext Markup Language) (RFC 1866) jako Internet draft pracovní skupinou IETF/IIR (Internet Engineering Task Force Internet Information Resources): dokumentový jazyk definovaný pomocí SGML užívaný na WWW.
- 9/1993: Mosaic pro PC, MAC a X-Window



## Výměna dokumentů na Internetu



- Identifikace dokumentů na Internetu: Uniform Resource Locator/Identifier/Name (URL/URI/URN[16]).
- URN: URI, kdy organizace zajišťuje trvalost odkazů (typicky překladovou službou).
- `metoda://server[:port]/cesta/soubor[#kotva]`
- Http, https, mailto, news, file, ftp, gopher, rlogin, telnet, tn3270, wais.
- Klient/server; http, https server.
- Klienti: Google Chrome, Firefox/Mozilla/Netscape Navigator, MS Internet Explorer, Opera, Galeon, Konqueror, Lynx, Mosaic, Amaya, HotJava, . . .
- Různé módy renderování v posledních verzích prohlížečů: zpětně kompatibilní (*quirk mode*) a standardy W3C dodržující (*standard mode*).
- Servery: Apache, Netscape Commerce, MS IIS, NCSA.

[16] http://www.w3.org/Addressing/#terms

## Značkovací jazyky na Internetu – vývoj



- 11/1993: Dave Raggett (HP) navrhuje HTML specifikaci s formuláři, tabulkami a rovnicemi
- 4/1994: HTML DTD test suite (Dan Connolly)
- 6/1994: MIT/CERN vytváří organizaci W3
- 6/1994: IETF vytváří pracovní skupinu HTML a HTML 2.0 specifikaci
- 11/1994: volně šířený prohlížeč Netscape 1.0
- 4/1995: Netscape Navigator 1.1
- 5/1995: Netscape a Sun se dohodli na podpoře Javy
- . . . World Wide Web Consortium (W3C).
- XML/XHTML, MathML.

☞ Úsečnost zápisu není důležitá.





## Dokumentové jazyky založené na SGML



☞ Nejrozšířenější je HTML.

☞ Vývoj HTML – různá *fixní* DTD (jedné třídy dokumentů).

☞ Koordinace W3C, velké nekompatibilní odchylky velkých firem, problémy s validací.

☞ DocBook DTD: svět Linuxu, dokumentace, knihy (O'Reilly, Kosek, Safari online).

☞ Další rozšířené dokumentové jazyky: TEI (Text Encoding Initiative) DTD.

☞ Rainbow DTD: formalizace RTF pro konverze z Wordu.

☞ WML (WAP), CALS, MATHML, TEXML, ...



## Co je to XML?



☞ Doporučení W3C[18]: rozšiřitelný značkovací *(meta)*jazyk.

☞ Univerzální a otevřený formát pro reprezentaci (a výměnu) téměř libovolné datové struktury – dat i dokumentů. Násobné a opakované použití jako u SGML.

☞ Celá sada technologií a formátů s XML souvisejících (XPointer, XLink, XSL) pro elektronickou výměnu dat (a dokumentů).

[18] http://www.w3.org/TR/REC-xml



## Proč XML? Desatero cílů.



XML (E**x**tensible **M**arkup **L**anguage) – zjednodušená verze SGML optimalizovaná pro použití na Internetu vyvíjená konsorciem W3C[17]. Desatero cílů:

☞ Přímočaré použití na Internetu.

☞ Široké spektrum použití/aplikací.

☞ Kompatibilita s SGML.

☞ Snadnost vytváření programů pro práci s XML.

☞ Absolutní minimum či absence volitelných rysů XML.

☞ Čitelnost a jasnost.

☞ Rychlost návrhu.

☞ Formální popis a návrh.

☞ Snadnost vytváření XML dokumentů.

[17] http://www.w3.org

## XML deklarace



```
<?xml version="1.0" encoding="iso-8859-2"
                            standalone="no"?>
```

- Musí být na prvním řádku dokumentu.
- Při `standalone` *yes* není třeba číst externí DTD.



## Vymezení XML



- Odlišení od HTML: rozšiřitelnost, možnost změn sémantiky značek. XML nespecifikuje ani sémantiku, ani množinu značek, umožňuje však značky definovat a definovat jejich strukturální závislosti. Sémantiku určují aplikace (webové prohlížeče jsou jedny z nich).
- Odlišení od SGML: zjednodušeně řečeno XML je SGML s restrikcemi. Odchylky jsou minimální (chování mezery).
- Technický úvod do XML Normana Walshe[19].

[19] http://www.xml.com/xml/



## Instrukce pro zpracování



Pokyny (procesní instrukce) pro specifické aplikace:

```
<?xml-stylesheet href="epd.css" type="text/css"?>
```

nebo

```
<?LaTeX \pagebreak?>
```

či

```
<datum>
 <?php echo Date "d.m.Y"?>
</datum>
```



## Syntaxe XML



- Slabší pojem než validita: **správná strukturovanost** (**well-formed**).
- Správně strukturovaný dokument by měla zpracovat každá XML aplikace.
- Validátory SGML umí obvykle i XML.
- Parsery: nsgmls, msxml, xerces, xmllint, ...
- Sekce CDATA pro pohodlnější psaní části dokumentů obsahujících významné znaky:

```
<moudrost>
<![CDATA[
 (1 < 2) & 2 = 1000 x 1
]]>
</moudrost>
```

- Nativní podpora MathML od Mozilly 1.1, Netscape 7.0. Renderovací stroj MathML pro MSIE: MathPlayer[22]. Techexplorer IBM[23]: plug-in pro Navigator a MSIE pro renderování TEXu, LATEXu a MathML.
- Možnosti výpočtů, validace, renderování na MathMLcentral[24].

[22] http://www.dessci.com/webmath/mathplayer/
[23] http://www.ibm.com/software/techexplorer/
[24] http://www.mathmlcentral.com





## XHTML? XHTML!

- Postupný přechod z HTML: XHTML[20], reformulace HTML jako modulární XML aplikace.
- XHTML čtou všechny XML-kompatibilní aplikace, a zároveň je možno psát **dobře zformované** (well-formed) dokumenty již nyní.
- Příklad XHTML dokumentu:

```
<!DOCTYPE html
  PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
  "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
  <html>
   <head>
    <title>simple document</title>
   </head>
   <body>
    <p>a simple paragraph</p>
   </body>
  </html>
```

[20] http://www.xhtml.org





## WML

- WML (Wireless Markup Language) je jazyk dokumentů, které jsou zobrazitelné na displejích mobilních telefonů: jsou dostupná DTD[25].
- Průmyslová asociace WAPFORUM[26]. (1997, Nokia, Ericsson, Motorola, Unwired Planet).
- WAP (Wireless Application Protocol) – komunikační protokol pro přenos informací z Internetu na mobilní zařízení. resp. `/packages/share/sgml-catalogs/WML`.
- WAP browsery: CCWAP[27], Nokia browser, WINWAP, ...
- Více paměti, silnější procesory: přechod na XHTML?

[25] http://www.wapforum.org/DTD/
[26] http://www.wapforum.org
[27] http://www.ccwap.com





## MathML: značkování matematiky na Webu

- MATHML[21] 2.0 (Mathematical Markup Language) – doporučení W3C. XML slovník/aplikace pro popis matematiky (struktury formulí i prezentačních forem).
- Příklad: kubickou křivku formalizovat pro další zpracování (vykreslení křivky, integrace, ...) na základě XML kódu:

```
<math xmlns='http://www.w3.org/1998/Math/MathML'>
  <msup>
    <mi>x</mi>
    <mn>3</mn>
  </msup>
</math>
```

- Použití jak pro sazbu, tak pro výpočet v systémech jako je Mathematica: přímočaré odlišení víceznačností: diferenciál $x$ od proměnné $\mathrm{dx}$ a součinu proměnných $dx$.

[21] http://www.w3.org/Math/

## Osnova



① Agenda (Typo Programming, úkoly do cvičení, bak. práce, web CSTUGu).

② ISO/IEC 10646, Unicode.

③ Formáty a normy související s XML (XLink, XPointer, RDF, XTM). AJAX. DOM.

④ Návrh (webový): úvod.

⑤ Webový design prakticky (Martin Kacvinsky).



## Příklad WML dokumentu



```
<?xml version="1.0"?>
<!DOCTYPE wml PUBLIC
 "-//WAPFORUM//DTD WML 1.1//EN"
 "http://www.wapforum.org/DTD/wml_1.1.xml">
 <wml>
   <card id="Card1" title="ccWAP WML ">
   <p>
   <!-- I am learning WML example -->
   I am learning the basics of WML.
   </p>
   </card>
 </wml>
```



## Technologie a formáty související s XML



- ISO/IEC 10646-1 Unicode.
- XLink, XPointer, XPath.
- RDF (Resource Description Framework) – metadata na WWW.
- CSS.
- XSL, XSL-FO.
- SVG.



## WML – pokračování



- Konfigurace http serveru pro WML: do souboru `~/.htaccess` je třeba přidat:

```
addtype text/vnd.wap.wml Wml
addtype Application/vnd.wap.wmlc Wmlc
addtype text/vnd.wap.wmlscript Wmls
addtype Application/vnd.wap.wmlscriptc Wmlsc
addtype image/vnd.wap.wbmp wbmp
```

- Prohlížení většinou pomocí apletů v prohlížeči, vývojová prostředí např. EasyPad Waptor[28].

[28] http://www.yourwap.com

## Unicode konsorcium, znaková sada Unicode



- **Unicode Consortium**[29] je nevýdělečná organizace, založená 1989 pro vývoj, rozšiřování a podporu používání Unicode Standardu. Členové konsorcia představují široké spektrum korporací a organizací v průmyslu počítačů a informačních technologií. Konsorcium je finančně podporováno pouze členskými poplatky.
- Respektuje ISO.
- Unicode: sada znaků specifikovaná konsorciem Unicode.
- Databáze znaků[30] poskytovaná konsorciem (k 1/2009 verze Unicode 5.1.0) či přehledné tabulky znaků[31].
- Znak s libovolným kódem z ISO 10646 můžeme zapsat pomocí entity `&#kod` (`kod` je v desítkové soustavě) nebo `&#xkod` (`kod` je šestnáctkově): code points $0$–$10FFF_{16}$.

[29] htpp://www.unicode.org/
[30] http://www.unicode.org/Public/UNIDATA/
[31] http:/www.unicode.org/charts/



## ISO/IEC 10646-1 standard



- Mezinárodní ISO standard ISO/IEC 10646-1, 1993: **Universal Multiple-Octet Coded Character Set** (UCS).
- První standard s ambiciózním cílem obsahovat všechny znaky používané v psaných jazycích, včetně matematických a dalších symbolů.
- Jednotné kódování, sekvence 16bitové (UCS-2, 65 536 pozic), nebo 31bitové (UCS-4, 2 147 483 648 pozic): proto **multi-octet**.
- Jazykově neutrální, zrušení předpokladů ekvivalence znaků a glyfů, neproporcionality hladkého textu.
- Kritériem přidání znaku je odlišnost ve významu i tvaru (CJK unifikace). Výjimky: Latin P, Greek P (capital rho), Cyrillic P (R).
- Struktura kódového prostoru UCS-2 (256 skupin po 256 buňkách), Basic Multilingual Plane (BMP).



- Jednoznačný trend použití pro rozšíření na WWW, v OS [cut & paste], značkovacích i programovacích jazycích.





- Privátní zóny pro privátní použití aplikací.
- Implementační úrovně 1–3; příklad dánského “A with ring above and acute accent”: `01FA` *reprezentace použitá v úrovních 1 a 2*: jeden *složený* znak; `00C5 0301` *“A with ring above” + combining acute accent* nebo `0041 030A 0301` *“A” + combining ring above + combining acute accent*: program implementující standard úrovně 3 musí být schopen zpracovat tyto *kompozitní sekvence* stejně jako výše uvedenou reprezentaci.

```
<odkaz xmlns:xlink="http://www.w3.org/1999/xlink"
  xlink:type="simple"
  xmlns:href="http://www.fi.muni.cz/~sojka/PB029/">
    Domovina PB029</odkaz>
```





## UCS a přenos znaků



- ☞ UCS-4 resp. UCS-2 “Universal Character Set coded in 4 resp. 2 octets”.
- ☞ UCS-2 pouze pro BMP (two-octet BMP form).
- ☞ Problémy s přímým přenosem textů v UCS (prostorově neefektivní) → transformační/kódovací metody UTF-8, UTF-7, UTF-16.
- ☞ UTF-8: znaky ASCII nahrazeny byty hex 00-7F. Ostatní znaky nahrazeny dvěma až šesti byty hex 80-FF.
- ☞ UTF-16 16bitové slovo obsahuje kód znaku, rozšíření: 1 znak zabere dvě 16bitová slova.
- ☞ UTF-7: pro SMTP protokol; kódovány i speciální znaky ASCII. Specifikovaný IETF, Internet Engineering Task Force, a formálně nesouvisející s ISO/IEC 10646. Výsledkem je text neobsahující kontrolní znaky a znaky nad `7F`.
- ☞ Detaily a přehled[32].

[32] http://www.nada.kth.se/i18n/ucs/unicode-iso10646-oview.html



## RDF



- ☞ RDF[35] (Resource Description Framework) – metadata na WWW.
- ☞ Model pro reprezentaci pojmenovaných vlastností a reprezentaci hodnot.
- ☞ Krok k sémantickému Webu?
- ☞ RDF validátor[36]:

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns="http://purl.org/dc/element/1.0/">
 <rdf:Description about="http://www.fi.muni.cz/~sojka/PB029/">
  <title>PB029 Elektronická příprava dokumentů</title>
  <description>Stránky předmětu PB029
    vyučovaného Petrem Sojkou na FI MU v Brně, CZ.</description>
  <author>Petr Sojka</author>
  <creator>Petr Sojka</creator>
  <date>2005-10-03</date>
  <format>text/html</format>
  <language>cs</language>
 </rdf:Description>
</rdf:RDF>
```

[35] http://www.w3.org/RDF/
[36] http://www.w3.org/RDF/Validator/



## XPointer a XLink



“Only connect! That was the whole of the sermon”
*E. M. Forster, (1879–1970)*

Doporučení XLink[33] **XML Linking Language** a draft XPointer[34] **XML Pointer language** pro specifikaci vazeb mezi XML dokumenty nebo jejich částmi.

- ☞ Zobecnění hypertextového odkazu známého z HTML, „zpětná kompatibilita“.
- ☞ Odkazy mezi více zdroji.
- ☞ Doplnění odkazů o metadata.
- ☞ Odkazy je možno uložit mimo odkazované dokumenty.
- ☞ Xpointer: přesné určení místa ve zdroji, případně výběr jeho části.
- ☞ Odkazy se vytvářejí pomocí elementů a atributů patřících do jmenného prostoru http://www.w3.org/1999/xlink.

[33] http://www.w3.org/TR/xlink/
[34] http://www.w3.org/TR/2001/WD-xptr-20010108/

## Agenda



- ☞ Princip: domácí úlohu do cvičení $n$ nosit i do dalších, nebo vystavit na stránce předmětu.
- ☞ Prémiové úlohy (záznamník body), Invex, bienále,. . . .
- ☞ Diskuzní fóra
- ☞ Témata BP a DP



## Závěrem k etapě o značkování



- ☞ Trendy: XML a navazující technologie, **Topic Maps** (ISO/IEC 13250) mapy témat, XTM 1.0[37]. Unicode.
- ☞ Výměna dokumentů/dat v heterogenním prostředí: B2B, B2C.
- ☞ Rozhraní a služby na Webu: SOAP, UDDI, WSDL, XLANG/XAML, XKMS.
- ☞ Ontologie: formální vyjádření sdílené konceptualizace daného oboru. Použití pro přesnější vyhledávání, Google.
- ☞ SMIL (Synchronized Multimedia Integration Language) pro interaktivní multimédia na Webu.
- ☞ Další zdroje, odkazy a literatura je na stránce předmětu (Kosek[38], xml.com[39], zvon.org[40], W3C[41], . . . ).

[37] http://www.topicmaps.org/xtm/1.0/
[38] http://www.kosek.cz
[39] http://www.xml.com
[40] http://www.zvon.org
[41] http://www.w3c.org/



## Kaskádové styly



"The Web is Ruined and I Ruined It![42]"
*David Siegel*[43]

- ☞ Informační web – výhody strikního oddělení obsahu a formy: **kaskádové styly**[44]. CSS, CSS2[45] (1998), CSS3[46].
- ☞ Jazyk stylů má formální gramatiku[47], proto je možná validace CSS[48].

[42] http://www.webreview.com/1997/04_11/webauthors/04_11_97_1.shtml
[43] http://www.dsiegel.com
[44] http://www.w3.org/Style/
[45] http://www.w3.org/TR/REC-CSS2/
[46] http://www.w3.org/TR/css3-roadmap/
[47] http://www.w3.org/TR/REC-CSS2/grammar.html
[48] http://jigsaw.w3.org/css-validator/



## Osnova dnešní přednášky



- ☞ Agenda
- ☞ Specifika návrhu webových stránek a publikování na Internetu
- ☞ Shrnutí webového designu
- ☞ Typografické tradice, cyklus přípravy dokumentů
- ☞ Základy návrhu publikací knižního typu

# Příklad CSS2

```
BODY              padding: 8px; line-height: 1.33
H1                font-size: 2em; margin: .67em 0
H2                font-size: 1.5em; margin: .83em 0
H3                font-size: 1.17em; margin: 1em 0
H4, P,
BLOCKQUOTE, UL,
FIELDSET, FORM,
OL, DL, DIR,
MENU              margin: 1.33em 0
H5                font-size: .83em; line-height: 1.17em; margin: 1.67em 0
H6                font-size: .67em; margin: 2.33em 0
H1, H2, H3, H4,
H5, H6, B,
STRONG            font-weight: bolder
BLOCKQUOTE        margin-left: 40px; margin-right: 40px


I, CITE, EM,
VAR, ADDRESS      font-style: italic
PRE, TT, CODE,
KBD, SAMP         font-family: monospace
PRE               white-space: pre
BIG               font-size: 1.17em
SMALL, SUB, SUP   font-size: .83em
SUB               vertical-align: sub
SUP               vertical-align: super
S, STRIKE, DEL    text-decoration: line-through
HR                border: 1px inset
OL, UL, DIR,
MENU, DD          margin-left: 40px
OL                list-style-type: decimal
OL UL, UL OL,
UL UL, OL OL      margin-top: 0; margin-bottom: 0
U, INS            text-decoration: underline
CENTER            text-align: center
BR:before         content: "\ A"
```





# Zpracování dokumentu s kaskádovými styly

- Parsing a vytvoření stromu dokumentu.
- Identifikace **výstupního zařízení**.
- Stažení všech stylů souvisejících s vybraným výstupním zařízením.
- Přiřazení hodnot všem **vlastnostem** výstupního zařízení, použitím pravidel **dědičnosti** a **kaskádování**.
- Pravidlo: **selektor** a **deklarace**.

```
BODY  font-size: 10pt
H1  font-size: 120%
```

- Kaskáda pravidel: autor (ve stránce), uživatel, uživatelův agent (nejnižší váha).
- Nejvyšší váhu má implicitně autor, uživatel ji ale může přebít (`!important`).





```
/* An example of style for HTML 4.0's ABBR/ACRONYM elements */

ABBR, ACRONYM     font-variant: small-caps; letter-spacing: 0.1em
A[href]           text-decoration: underline
:focus            outline: thin dotted invert

@media print
  @page           margin: 10%
  H1, H2, H3,
  H4, H5, H6      page-break-after: avoid; page-break-inside: avoid
  BLOCKQUOTE,
  PRE             page-break-inside: avoid
  UL, OL, DL      page-break-before: avoid



@media speech
  H1, H2, H3,
  H4, H5, H6      voice-family: paul, male; stress: 20; richness: 90
  H1              pitch: x-low; pitch-range: 90
  H2              pitch: x-low; pitch-range: 80
  H3              pitch: low; pitch-range: 70
  H4              pitch: medium; pitch-range: 60
  H5              pitch: medium; pitch-range: 50
  H6              pitch: medium; pitch-range: 40
  LI, DT, DD      pitch: medium; richness: 60
  DT              stress: 80
  PRE, CODE, TT   pitch: medium; pitch-range: 0; stress: 0; richness: 80
  EM              pitch: medium; pitch-range: 60; stress: 60; richness: 50
  STRONG          pitch: medium; pitch-range: 60; stress: 90; richness: 90
  DFN             pitch: high; pitch-range: 60; stress: 60
  S, STRIKE       richness: 0
  I               pitch: medium; pitch-range: 60; stress: 60; richness: 50
  B               pitch: medium; pitch-range: 60; stress: 90; richness: 90
  U               richness: 0
  A:link          voice-family: harry, male
  A:visited       voice-family: betty, female
  A:active        voice-family: betty, female; pitch-range: 80; pitch: x-high
```





# Vytváření kaskádového stylu

- Koskův článek o dynamickém HTML[49].
- Kuchařka Dave Raggetta[50].
- Příklady webů pánů Staníčka, Pecháčka[51], Řezáče[52] či Kacvinského.

[49] http://www.kosek.cz/clanky/dhtml/styly.html
[50] http://www.w3.org/MarkUp/Guide/Style
[51] http://www.y3k.cz/epd/
[52] http://www.fi.muni.cz/

## Dynamické HTML



- ☞ Skriptování na straně klienta (JavaScript, standard ECMAScript; VBScript).
- ☞ Skriptování na straně serveru. CGI skripty (Perl, PHP).
- ☞ API: Objektový model dokumentu W3C DOM.
- ☞ DHTML vs. Java.[54]
- ☞ Sada technologií AJAX (Asynchronous JavaScript and XML).

[54] http://www.kosek.cz/clanky/dhtml/index.html



## XSL, XSLT, XSL-FO



"Use CSS when you can, use XSL when you must."
*W3C*

- ☞ Proč XSL? CSS nebo XSL[53]?
- ☞ XSLT – transformační jazyk XSL.
- ☞ XSL-FO – formátovací objekty XSL.

[53] http://www.w3.org/Style/CSS-vs-XSL



## WWW design



„Já věřím tomu, co je vytištěno černou barvou na papíře. Média jsou pro okultisty."
*František Štorm, 1. střešovická písmolijna*

- ☞ Design dokumentu je *závislý* na přenosovém médiu (pravidla pro knižní design pro WWW design vesměs *neplatí*). Nutno proto mít *různé* verze dokumentů pro různá média.
- ☞ Specifika WWW: čtení na obrazovce počítače (poměr stran, rozlišení), různé „oči" (prohlížeče), různá „viditelnost/osvětlení" (tloušťka/rychlost linky síťového připojení), snadnost vyjádření struktury, vazeb (hypertext, nelineárnost).
- ☞ *Informační* × *prezentační* web.

## Desatero WWW designu



- ☞ Thou shalt put no page element before Content, for Content is thy lord and It is a jealous lord.
- ☞ Thou shalt make thy webpages to load quickly.
- ☞ Thou shalt respect the trademarks, copyrights, and reputation of thy neighbor.
- ☞ Thou shalt use graphics and animations that are relevant and meaningful, and thou shalt not scatter them carelessly as birds scatter seeds on hard ground.
- ☞ Thou shalt not resize thy neighbor's browser window nor shall thou disrupt the function of thy neighbor's "back" button nor shall thou force text to a size that is unreadably small on thy neighbor's 1280x1024 display.
- ☞ If thou art inspired to create a webpage of lasting value, then thou shalt make it in such a manner so that thy neighbor may bookmark or print it. Twould be even better if he could do both.



## Kuchařka dobrého WWW designu



- ☞ Web není WYSIWYG (návrhář nezná předem čtenářova nastavení).
- ☞ Klíčem k úspěchu je originální obsah.
- ☞ Vytváření stránek se *liší* od zpracování textu.
- ☞ HTML/XHTML/XML je *strukturální* jazyk.
- ☞ Ne každý uživatel má grafického klienta nebo funkční zrak.
- ☞ Většina skvělých triků není až tak skvělá (zmenšují čtenářskou obec).
- ☞ Důležité je časování (čtenář a jeho konektivita se mění, špičkové věci rychle zastarávají).
- ☞ Dodržování standardů – statistika není argument pro nekorektní stránky.
- ☞ Uživatelé mají rádi rychlé stránky.



- ☞ Thou shalt make the elements which must be seen and comprehended first so that they load first.
- ☞ Thou shalt use text for text and graphics for graphics and never graphics where text is appropriate, under threat of banishment to the pit of darkness where searches and indexes are unknown.
- ☞ Thou shalt make thy pages to be both standards compliant and backwards compatible, but where backward compatibility shall cause the page to load slowly, thou shalt leave those with weak and old browsers in the wilderness they have chosen for their home.
- ☞ Thou shalt not steal bandwidth by including images or other objects housed on thy neighbor's website on thy own pages, nor shalt thou place thy neighbor's website in a frame on thy page. In all things remember to treat thy neighbor's website as thou wouldst have him or her treat thine.



- ☞ Kontrolujte se.

## Budoucnost WWW (designu)



- ???
- Různá kódování (češtiny) $\longrightarrow$ Unicode.
- Standardy W3C: SVG, XML/RDF, XLink, Xpointer, XTM, ...
- Závislosti časové (SMIL?).
- Závislosti na výstupním zařízení (PDA a WAP [$640 \times 260$], multimédia). XSL? XSL!
- Dynamické stránky (Java, Flash, CGI skripty, výhody a nevýhody).
- DOM (Document Object Model).
- Napojení na databáze, PHP, ASP.
- VRML (Virtual Reality Markup Language), multimodalita.



## Chyby návrhu WWW stránek



- Chybějící navigační podpora.
- Nekonzistence (nejednotnost) a přeplácanost.
- Příliš pomalá doprava.
- Osiřelé stránky.
- Dlouhé, těžké stránky.
- Zastaralé informace.
- *Nad*užívání nejmodernějších technologií.
- Pohyb bez příčiny.
- Složitá URL.
- Nestandardní barvy odkazů.
- Používání rámů[55].

[55] http://www.useit.com/alertbox/9612.html



- Velikost dat (obrázky, rozlišení), rychlost linek.
- Worth world wide web: Dogma W4[61]?

[61] http://www.pixy.cz/dogmaw41/



## Odkazy k technologiím na webu



- Weby W3C[56], OASIS[57].
- V češtině například Koskovy články[58], W3Schools[59] nebo tutoriály na Živě[60].
- Knihy Jacoba Nielsena, Davida Siegela, Pavla Satrapy či Petra Staníčka.

[56] http://www.w3c.com
[57] http://www.oasis-open.org/cover/sgml-xml.html
[58] http://www.kosek.cz
[59] http://www.w3schools.com/
[60] http://www.zive.cz/H/Developerdaily/

## Jednotnost (makrotypografie)



- ☞ Pravidelné šedivé mřížky (citlivost oka).
- ☞ Nezbytnost bílých míst (horror vacui), nesou informaci (oddělují, spojují)!
- ☞ Dvoustránka (double-spread), vyváženost, příklad výjimky: zkrácení dvoustránky.
- ☞ Jednotnost umístění stránky, prvků ve stránce.
- ☞ Řádkový rejstřík (protisk!).



## Abstrakt



V první části přednášky budou zmíněny základní zásady často zcela absentující etapy přípravy dokumentů, a to návrhu (knižních) dokumentů. Jádrem přednášky pak bude přehled typografických pravidel české sazby, které se ustálily tradicí v průběhu minulých staletí, ale dnes příliš často ctěny nejsou, zvláště v publikacích vznikajících na stole (DTP) příležitostných sazečů.

„Rozhodně, prosím, opusťte vody nepoučených „nalévačů textu“ či „myšařů“ a začněte vyrábět sazbu, již je nejen radost číst, ale na níž i oko se zalíbením spočinouti může.“
*Michal Ptáček, 1995 (seriál v PC World)*



## Osnova dnešní přednášky



- ☞ Základy návrhu publikací knižního typu.
- ☞ Sazba, základní pojmy, písma.
- ☞ Typografická pravidla pro české texty.
- ☞ Principy návrhu knih.
- ☞ Sazba, sázecí systémy.
- ☞ Základy typografie.



## Principy návrhu knih



“No, Watson, this was not done by accident, but by design.”
*Sherlock Holmes*

- ☞ Generace návrhářů, typografů samouků; samozvaných expertů.
- ☞ Nejčastější chyba: instantní řešení, *absence* **návrhu** (**designu**).
- ☞ Schopnost odlišovat dobrý návrh od špatného.
- ☞ Černobílost pohledu, protikladnost požadavků, výjimky. „Šaty dělaj’ člověka“.
- ☞ Knižní design: co je kniha; specifika?
  - ❑ Poctivost obalu (na rozdíl od časopisu).
  - ❑ Vazba po blocích (na rozdíl od vazby na stříšku).
- ☞ Forma
  - ❑ Jednotnost.
  - ❑ Přenos informace (raison d’être).
  - ❑ Struktura (rejstřík, obsah).

## Struktura



- ☞ Obsah, seznamy tabulek, obrázků.
- ☞ Křížové odkazy, jednoúrovňové, dvojúrovňové (nepřímé).
- ☞ Rejstřík.
- ☞ Paginace, paginace vakátů.
- ☞ Spuštěné číslování na úvodních stránkách kapitol.
- ☞ Živá záhlaví, zápatí, co do nich (která úroveň nadpisů)?
- ☞ Zkracování nadpisů do záhlaví, číslování.



## Jednotnost (mikrotypografie)



- ☞ Mezipísmenný proklad.
- ☞ Mezerování (data, zkratky, sazba telefonních čísel, mezivětné mezery).
- ☞ Zavěšená interpunkce.
- ☞ Odstavcová odrážka.
- ☞ „Zákon malých celých čísel“.



## Knižní design – praktické rady



Stránka/dvoustránka – maketa

- ☞ Rozměry stránky: standardní velikosti (řada A, B, kvarto, folio, . . . ).
- ☞ Aspect ratio: poměr šířky k výšce; důvody:
  - ❑ stabilnost držení v ruce (těžiště blíže než délka dlaně);
  - ❑ počet znaků na řádku 40–70.
- ☞ Okraje: 1:6:2, 1:8:3, ořez, odpad (montáž), vazba, avantgardní design, marginálie jako prvek designu.
- ☞ Umístění stránky na listu: světlost záhlaví, zápatí.
- ☞ Symetrie a asymetrie.
- ☞ Mřížkový design.



## Přenos informace



“Printing should be invisible.”
*M. Warde*

- ☞ Spojitost nadpisu s následným textem.
- ☞ Spojitost popisky s tabulkou, obrázkem.
- ☞ Způsoby vyjádření vazeb, spojitostí, nebezpečí (duktus/velikost).
- ☞ Patkové, bezpatkové písmo. Střídmost.
- ☞ Čitelnost. Počet znaků na řádek.
- ☞ Velikosti stránek. Souhrn standardů – paskvil.
- ☞ Volba písem – střední výška písma.

**xii.** vakát

**1** první kapitola

**. . .**

**[tiráž]**





## Písmo a rozpal



☞ Písmo, základní parametr návrhu.

- ❑ Konotace – klasická (Antikva) × avantgardní (Grotesk) písma.
- ❑ Čitelnost (patkové × bezpatkové).
- ❑ Velikost (mladí × staří čtenáři).
- ❑ Rodina (Times × Palatino), anamorfní zmenšení písma (na 24/25 ve vertikálním směru) v případě nutného přizpůsobení Computer Modern (Monotype 8a), závisí na výstupním zařízení (kontrast mezi tenkými a silnými liniemi).
- ❑ Základní velikost mezislovní mezery (řeky, řešení východů kapitol).
- ❑ Jednotnost: řezy téže rodiny, i pro matematiku (zatím malý výběr).

☞ Rozpal:

- ❑ 10/12, světlost, sazba na řádek;
- ❑ typ textů (diplomová práce, inzerát).



## Dobrý návrh nekřičí



“Good typography is not loud. A resonably applied and readable type style is the first ingredient of good composition. Today, asymmetric composition is certainly feasible, however, typography is intrinsically symmetrical. A text block without indents is unclear. Indents help the reader by reinforcing the logical order of the text. The best typography is invisible to the reader and serves to transmit the thoughts and intent of the author. Beautiful text, a text well-composed, is legible. One of the highest virtues of good typography is its subtle elegance. It is not the duty of the typographer to consciously display or emulate the style of current trends, nor to reflect the spirit of the times. Typography must be itself, it must be pleasing to the eye and not tiring. Good typography has absolutely nothing to do with remarkable or exotic type styles. This is the opinion of amateurs. The essence of letterform is not modernity, but readability.”

Jan Tschichold



## Číslování a řazení v knize



**i.** patitul

**ii.** vakát

**iii.** titul

**iv.** katalogizace publikace, copyright, [tiráž]

**v.** předmluva k vydání

**vi.** obecná předmluva

**vii.** ditto, pokračování

**viii.** vakát

**ix.** obsah

**x.** ditto, pokračování

**xi.** slovníček

## Základní pravidla sazby – pokračování, osnova



- ☞ Odstavec, odstavcová odrážka.
- ☞ Zarovnaný text.
- ☞ Nezarovnaný text.
- ☞ Dělení slov.
- ☞ Znaménka.
- ☞ Korekturní znaménka
- ☞ Pro detaily předchozích přednášek viz Beranův Aktualizovaný typografický manuál[63], příslušné ČSN, Ptáčkův text v PC Worldu 1994.

[63] http://www.dtpstudio.cz/obchod/promoakce/typomanual.htm



## Základní typografické pojmy – osnova



- ☞ Tradice, omezující podmínky, v designu explicitně nespecifikované.
- ☞ Normy (ON 88 2503,...) a jejich respektování. http://www.typotext.cz/radce7b_1.html[62]
- ☞ Znaky.
- ☞ Písmová osnova.
- ☞ Kresba písmového znaku.
- ☞ Měření písma.
- ☞ Typ písma.
- ☞ Rodina písma.
- ☞ Klasifikace tiskových písem.

[62] http://www.typotext.cz/radce7b_1.html



## Osnova dnešní přednášky



- ☞ Sazba, sázecí systémy. Co je TEX?
- ☞ Vznik a vývoj TEXu.
- ☞ Základní principy, řádkový a stránkový zlom.
- ☞ Základy digitální typografie, užití pro elektronické publikování. Pravidla sazby.
- ☞ Vše, co jste chtěli vědět o TEXu, ale báli jste se zeptat.





- ☞ Fonty.
- ☞ Mezislovní mezera.
- ☞ Vyplňování řádků.
- ☞ Řádkový proklad.

## Digitální revoluce a její důsledky pro černé řemeslo



- Digitální revoluce koncem sedmdesátých let.

"Mathematics books and journals do not look as beautiful as they used to."
*Donald Ervin Knuth*[64]

- Počítač: od imitace psacího stroje po *sázecí* systém DTP.
- Typografie matematiky, matematika/informatika slouží typografii: místo metalurgie *informatika*, matematický popis křivek (kubické splajny v programu pro návrh písem METAFONT).

"Go forth and create *masterpieces of electronic publishing art.*"
*(paraphrase of Donald Knuth [TEXbook, p. 303])*

[64] http://www-cs-faculty.stanford.edu/ knuth/



## Motivace



„Dokumenty v TEXu jsou mstou akademiků studentům, takže přežijí do té doby, dokud budou přežívat zatvrzelci v akademických sborech (řádově několik tisíciletí)."
*Jiří Hlavenka, (1964– )*

:-( Naproti tomu:
„Rozhodně, prosím, opusťte vody nepoučených ‚nalévačů textu' či ‚myšařů' a začněte vyrábět sazbu, již je nejen radost číst, ale na níž i oko se zalíbením spočinouti může."
*Michal Ptáček*



## Sázecí systémy



"GUIs normally make it simple to accomplish simple actions and impossible to accomplish complex actions."
*Doug Gwyn*

- **Sazba**: realizace návrhu nad značkovanými daty.
- WYSIWYG (WYSIAWYG): InDesign, Pagemaker (Adobe[65]), QuarkXpress (Quark[66]), 3B2 (Advent Publishing[67]).
- Dávkové systémy: TEX[68], troff/groff/nroff/runoff, Lout[69] (nutnost kontroly návrhu dokumentů ve *finální* podobě, ne jen na obrazovce).
- Textové procesory Word, AmiPro, ... sem de facto nepatří (nedovedou některé docela zásadní potřeby sazeče – nedělitelná roztažitelná mezera, fixace zlomu pro různá výstupní zařízení apod.).

[65] http://www.adobe.com
[66] http://www.quark.com
[67] http://www.3b2.com
[68] http://www.cstug.cz/
[69] http://snark.ptc.spbu.ru/ uwe/lout/



## Typografické tradice – řemeslo a umění



- „Dnes, kdy je typografie neprávem potlačována a kdy se typografem cítí každý, kdo dokáže vytvořit na počítači stránku textu, si vám dovolíme přiblížit dobu, kdy typografie byla řemeslem a uměním."
- „Budiž světlo!"
- Učební doba čtyřletá až pětiletá.
- Po staletí používané zvyky se staly tradiční, vytvoření národních typografických zvyklostí. (Vojtěch Preissig 1873–1944, Oldřich Menhart 1897–1962, Oldřich Hlavsa 1909–, František Muzika), Muzeum knihy ve Žďáru nad Sázavou.

"Those old guys stole all our best ideas."
*Frederick Goudy*

- ☞ Volně šiřitelný (vývoj hrazen granty).
- ☞ S výstupem *nezávislým* na výstupním zařízení.
- ☞ Jednoduchý základ pro sazbu: model box, glue, penalty.
- ☞ “A computer program of which a professor of computer science might be proud of.” (DEK)
- ☞ Optimalizovaný, vysoce efektivní, využívající nejrychlejší algoritmy své doby – při vývoji nalezeny nové informatické metody a datové struktury (trie).





## Programování sazby – TEX

- ☞ Algoritmy počítačové sazby v sázecím systému TEX
  1. Zlom řádků.
  2. Zlom stránky.
  3. Dělení slov.
  4. Umisťování obrázků, viditelnost.
  5. Umisťování poznámek pod čarou.
  6. Sazba matematických výrazů:

$$\sqrt{\left(\int_0^\infty \sqrt{\frac{x^2}{2}+1}\right)}$$

- ☞ `$$\sqrt{\left(\int_0^\infty \sqrt{\frac{x^2}{2}+1}\right)}$$`
- ☞ (Makro)programování sazby, otevřenost systému.
- ☞ OSS, CSTUG[70], CTAN[71], pros & cons.

[70] http://www.cstug.cz/
[71] http://www.cstug.cz/ctan?





## Co TEX není?

- ❑ Editor.
- ❑ Program na grafiku (na to slouží komplementární program METAFONT přibližně stejné velikosti).
- ❑ WYSIWYG (nadstavby jako LyX existují).
- ❑ Rychle naučitelný (strmější učicí křivka).





## Co je TEX?

- ☞ Sázecí autorský systém.
- ☞ Programovatelný, t.j. s vlastním makrojazykem (s vyjadřovací silou Turingova stroje – byl v něm například pro zábavu implementován interpret jazyka BASIC); výhoda pro cca 2 % populace.
- ☞ Dávkový: ze vstupu `$c=\sqrt{a^2+b^2}$` dostaneme $c = \sqrt{a^2 + b^2}$: http://tex.mendelu.cz[72], http://sciencesoft.at/index.jsp?link=latex[73].
- ☞ Otevřený/rozšiřitelný.
- ☞ Portabilní (od Atari či dvoudisketového PC XT po Cray).
- ☞ Stabilní ($256 za nalezení chyby), verze $\longrightarrow \pi$.
- ☞ Dobře dokumentovaný (vyšel knižně).

[72] ukázka 1
[73] ukázka 2

## TEX dnes I: PDF



- Nový formát pro elektronické knihy fy Adobe.
- Otevřenost TEXu: pdfTEX[76] Hàn Thế Thành.
- Kalkulačka v PDF (+JavaScript): dříve než Adobe.
- AcroTEX[77].
- PDF není až tak portabilní jak říká název – místo pro DVI3?

[76] http://www.cstug.cz/pdftex/
[77] http://www.math.uaktorn.edu/~dpstory/acrotex.html



## Vznik a vývoj TEXu



- 1977, korektura *The Art of Computer Programming*.
- 1978, první verze; TEX82, METAFONT84; osmibitový TEX 3 (1990); rozšiřování o výstupní formáty (PostScript).
- 1992: zmrazení dalšího vývoje, pevný bod, jen opravy chyb.
- Vznik makrobalíků.
- Téměř žádný marketing, uživatele sdružuje TUG a lokální sdružení uživatelů – LUG.



## TEX dnes II: Internet (HTML/XML/MATHML)



- HyperTEX (Los Alamos archiv[78]).
- LATEX2HTML[79], konvertor z nejznámějšího makrobalíku do HTML v Perlu.
- TEX4ht, nebo jiné mnohé komerční implementace TEXu s alternativním výstupem do HTML jako TEXpider[80] (MicroPress).
- eeTEX (pro přímou sazbu SGML).
- TTH[81]: konvertor TEXu do HTML.
- IBM techplorer Hypermedia Browser[82] – plug-in do Netscape a Internet Exploreru.
- MathType[83] pro psaní matematiky ve Windows.

[78] http://xxx.lanl.gov
[79] http://www-texdev.mpce.mq.edu.au/l2h/mathdocs/mathdocs/
[80] http://www.micropress-inc.com/webb/wbstart.htm
[81] http://hutchinson.belmont.ma.us/tth/
[82] http://www-4.ibm.com/software/network/techexplorer/
[83] http://www.mathtype.com



## TEX dnes: pro klasickou publikační činnost



- Zejména pro matematiku a všude tam, kde je možná algoritmizace zpracování ($$).
- Velká nakladatelství technické literatury a časopisů Springer-Verlag, Elsevier Publishers, Kluwer sází v TEXu časopisy, sborníky.
- Databázové publikování: Dopravní podnik města Brna (tabulky jízdních řádů pro zastávky, řidiče, dispečery), rozvrhy FI MU, studijní program FI MU přímo z databáze informačního systému MU.
- Slovníky (LEDA[74]) a první díl encyklopedie (Diderot[75]).
- Jádro TEXu či jeho algoritmy v sázecích systémech 3B2, InDesign, troff, Lout či v textovém procesoru Word.

[74] http://www.leda.cz
[75] http://www.diderot.cz

## Perspektivy TEXu a jeho následníků v tomto tisíciletí



- TEX pevný bod v chaosu postmoderního světa?
- Interaktivní dokumenty: otevřenost umožňuje rychle reagovat na nové formáty a standardy.
- Rozšiřování požadavků na **autorské** systémy.
- Bude se opakovat historie s tím, že v TEXové komunitě navržené a vyzkoumané postupy, algoritmy a přístupy se stanou základem většiny komerčních aplikací v oblasti DTP?
- embedded systémy [části] TEXu součástí větších projektů.





- TEXML[84] (Doug Lovell, IBM): XML reprezentace TEXové syntaxe.
- WebEQ[85]: sada programů včetně Javovského apletu pro rastrování matematiky.
- Texi2html[86]: konverze Texinfo formátu do html.
- Word2TeX[87]: konfigurovatelný konvertor z rozšířeného měnícího se formátu do stabilního.
- ASTER[88] systém pro čtení LATEXového zápisu a dokumentů pro nevidomé
- a mnoho dalších . . .

[84] http://www.alphaWorks.ibm.com/formula/texml/
[85] http://www.webeq.com/
[86] http://www.mathematik.uni-kl.de/ obachman/Texi2html/
[87] http://www.word2tex.com/
[88] http://www.cs.cornell.edu/Info/People/raman/aster/demo.html



## Jak se na to připravit?



- Instalovat si TEX a učit se ho:
- učit se, učit se, učit se.
- Stát se členem sdružení CSTUG[94] a účastnit se aktivně jeho akcí a projektů.
- Úcta k tradicím, osvěta.
- Další studium: www.sazba.cz[95] (Kapr: Sto a jedna věta o knižní úpravě), www.typo.cz[96] (písma, PostScript, . . . ), Beranův Typografický manuál[97].

> "A man who would letterspace lower case would steal sheep."
> *Frederick Goudy, 1894–1945*

> "If this wisdom needs updating, it is chiefly to add that *A woman who would . . .*"
> *Robert Bringhurst*

[94] http://www.cstug.cz
[95] href://www.sazba.cz
[96] href://www.typo.cz
[97] http://www.dtpstudio.cz/obchod/promoakce/typomanual.htm



## TEX dnes III: Java, Unicode, OpenType



- NTS: reimplementace TEXu bez optimalizací v Javě (Karel Skoupý, FI MU, nyní ÚVT MU), a následný výzkum.
- PDF prohlížeč v Javě[89] → vysoce portabilní prvotřídně vysazené dokumenty s minimem přenosu dat v dohledu.
- Omega[90]: mnohem více než šestnáctibitový TEX; konverze LATEXu do MathML.
- PATLIB[91], opatgen, upatgen, . . .
- XeTEX[92], viz sobotní přednáška CSTUG[93].

[89] http://www.adobe.com/products/acrviewer/main.html
[90] http://omega.cse.unsw.edu.au:8080/index.html
[91] http://www.fi.muni.cz/~xantos/PATLIB/
[92] http://scripts.sil.org/xetex
[93] http://bulletin.cstug.cz/pozvanka/

## Algoritmy řádkového a stránkového zlomu



- Algoritmus First fit: sazba po řádcích, výběr prvního možného zlomu. Nenáročné na výpočetní kapacitu.
- Algoritmus Best fit: zlom po řádcích, výběr nejlepšího zlomu na řádku bez přihlédnutí k předchozím možnostem. Míra optimality řádku je měřena mírou odchylky od ideální velikosti mezislovní mezery. V TEXu je míra škaredosti (badness) $100 \cdot |r|^3$, kde $r$ je poměr rozdílu přirozené a požadované délky řádku ku součtu roztažitelných a stlačitelných složek lepidla (glue).
- Algoritmus Optimum fit: optimalizace zlomu v rámci celého odstavce naráz s využitím modelu 'box, glue a penalty'. Úloha podobná hledání nejkratší cesty v grafu, každé cestě odpovídá škaredost, která se minimalizuje. Stačí zvažovat pouze možnosti hezčí než zadaná tolerance.



## Základní pravidla sazby – pokračování, osnova



- Čísla a zkratky.
- Hladká sazba.
- Smíšená sazba.
- Korekturní znaménka.
- Typografické pravítko.
- Příprava textů, rukopis.
- Sazební obrazec.
- Formáty papíru, počítání rozsahu sazby.
- Korekturní znaménka.
- Digitální formáty písem



## Parametry algoritmu řádkového zlomu TEXu



- Až tři průchody: první bez dělení slov, limit škaredosti je `\pretolerance`. Pak jsou rozdělena všechna slova v odstavci a proveden druhý pokus s limitem škaredosti `\tolerance`.
- Pokud je nastaven registr `\emergencystretch` na kladnou délku (zvětšení roztažitelnosti mezislovní mezery) a jde o TEX3.0 či vyšší, provede se třetí pokus o zlom s těmito limity roztažitelností.
- Další parametry ovliviňující algoritmy zlomu: řádkového: `\adjdemerits`, `\linepenalty`, `\hyphenpenalty`, `\exhyphenpenalty`, `\looseness` a stránkového: `\interlinepenalty`, `\clubpenalty`, `\widowpenalty`, `\brokenpenalty`,...
- Ladění parametrů zlomu pro daný projekt (Diderot).
- Detailní český popis variant algoritmů zlomu je k nalezení v DP FI MU Jana Pazdziory.



## Digitální formáty písem



- Formáty METAFONT, PostScript a TrueType (OpenType).
- Princip oddělení metrické informace potřebné při sazbě od informace o tvarech znaků.
- Metrické informace: rozměry jednotlivých znaků, mezipísmenný kerning, ligační programy. Formáty TFM a AFM, program afm2tfm.
- Tvary znaků jsou popsány jako bitmapy (optimalizace znaků pro obrazovku, PostScript Type 3) nebo pomocí křivek obrysů či per. PostScript Type1 používá pro popis obrysů kubické Bézierovy křivky, formát TrueType výpočetně méně náročné kvadratické Bézierovy křivky a METAFONT parametrická pera.
- Detailní český popis digitálních formátů písem je k nalezení v DP PřF MU Pavla Janíka a DP FI MU Jana Vidy.

## Principy makroprogramování



- Makrojazyk TEXu: složité sázecí postupy skládané z 300 základních příkazů (primitiv).
- (Ne)vhodnost makroprogramování (versus detailní značkování), minimalizace vstupu.
- Input procesor: řádky nezávislé na operačním systému.
- Token procesor.
- Expand procesor.
- Hlavní procesor.
- Detaily v češtině jsou v Olšákově TeXbooku naruby (TBN).



## *hz*-algoritmus



- Motivace: Gutenbergova 42řádková Bible měla 290 písmen abecedy proměnné šířky, stejná mezislovní mezera při sazbě do bloku.
- Principy: kerning, scaling, composing.
- URW, Karrow, Zapf, neúspěch (fonty).
- Implementace v pdfTEXu (Thành).
- Implementace v InDesignu (Karrow).
- Detailní český popis modifikací zlomového algoritmu písem je k nalezení v DP Mirky Misákové a dizertaci Hàn Thế Thành.



## Makrobalíky



- Znovuobjevování kola většinou není potřeba: tisíce balíků maker na CTAN (Comprehensive TEX Archive Network).
- Katalog balíků na CTAN.
- Příklady balíků, kniha LATEX Companion.
- CD/DVD s distribucí TEXlive 2007 založenou na web2c.



## Implementace algoritmu dělení slov



- Dizertace a program `PatGen` Franka Lianga.
- Princip soutěživých vzorů, nezávislost na jazyce.
- Vytváření a generování vzorů.
- `OPatGen`, knihovna PatLib, upatgen.
- Detailní popis algoritmu dělení slov je k nalezení v DP FI MU Davida Antoše.

## PostScript – vlastnosti jazyka



- Nezávislost na zařízení (různé rozlišovací schopnosti výstupních zařízení).
- Souřadné systémy, prostor uživatele, prostor zařízení.
- Interpreter – PostScriptový RIP, hardwarový nebo softwarový (Ghostscript, Ghostview).
- Zásobníkový programovací jazyk, postfixový zápis v reverzní polské notaci.
- Podpora různých barevných prostorů (CMYK, RGB), direktní barvy.
- Nezávislost na platformě – vytváření/generování PostScriptu aplikacemi.



## InDesign



- Komerční program firmy Adobe pro WYSIWYG počítačovou sazbu.
- Zlomový algoritmus (převzat z TEXu) s pomocí Davida Karrowa (URW).
- Škálování znaků.
- Optický proklad (David Kindersley: Optical letter spacing for new printing systems, 1976).
- Příklady zlomu a chyb.[98]

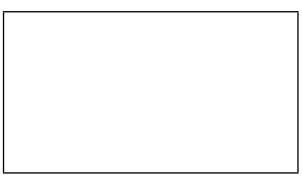
[98] http://www.fi.muni.cz/~sojka/PB029/xznamen/



## Postscript – příklad



```
 1  %!PS-Adobe-2.0
 2  %%Title: Priklad Postscript programu
 3  /Helvetica-Bold findfont
 4  50 scalefont setfont
 5  0 0 moveto 200 0 lineto
 6  200 100 lineto 0 100 lineto
 7  closepath 0.8 setgray fill
 8  0 setgray
 9  30 40 moveto
10  (FI MU)show
11  showpage
```

- 



## Jazyky pro popis stránky



- PDL – jazyk pro popis stránky, kombinace textu s grafikou.
- Formát firmy Adobe (Geschke, Warnock, 1982, Apple Laserwriter, Aldus Pagemaker), nyní de facto standard.
- Linotronic 1986, svět osvitových jednotek (vysoké rozlišení).
- Vývoj/podpora – Level1, Level2, (Level3).
- Návazné/odvozené/související produkty/standardy: PDF, písma Type1 (vektorový popis), Type3 (bitmapy).
- Nově OpenType[99].

[99] http://www.microsoft.com/typography/developers/opentype/default.htm

## PDF – vlastnosti



- *Proprietární* formát firmy Adobe.
- Objektový *formát*, není to programovací jazyk.
- Portabilita je dlužná svému jménu `:-(`. Neportabilní záložky (závisí na systémovém fontu).
- Binární formát, nelze jednoduše editovat (tabulka xref na konci PDF odkazuje na jednotlivé byty souboru).
- Struktura PDF, základní objekty, struktura souboru, popis stránky, komprese.
- Srovnání s DVI, PS.
- Možnost optimalizovat pro web, možnost vkládat JavaScript přímo do PDF.



## Struktura PostScriptového souboru



- Záhlaví, komentář Bounding Box, úrovně PostScriptu (Level).
- Definice procedur.
- Nastavení dokumentu.
- Jednotlivé stránky.
- Konec dokumentu.



## Způsoby vytváření PDF



- pdfTEX!
- ghostscript: dávka ps2pdf.
- Adobe Acrobat Distiller.
- PDF Writer (virtuální tiskárna, minimální optimalizace).
- Adobe Acrobat Capture (OCR).
- PStill, . . .



## PostScriptové objekty a jejich zpracování



- Prostor uživatele a prostor zařízení.
- Skládání afinních transformací (CTM, operátory pro modifikaci aktuální matice transformace: rotace, translace, zmenšení/zvětšení).
- Cesty.
- Konstrukce Bézierovy (kubické) křivky.
- Detaily v češtině jsou ve Zpravodaji CSTUG 1–3/2001[100].

[100] http://bulletin.cstug.cz/pdf/bul_0013.pdf

## Barevná separace a barevné prostory



- Konverze RGB na CMYK (azurová, purpurová, žlutá, černá).
- UCR: under color removal.
- Separace složek, přímé (direktní) barvy.



## Verze PDF



- Úrovně jazyka 1.0 až 1.7 pro Acrobat 1 až Acrobat 8. Aktuální specifikace PDF 1.7 má přes tisíc stran.
- PDF-X: Ořezaná verze jazyka podporovaná ANSI, ISO a firmami v předtiskovém byznysu[101].
- PDF-X1, PDF-X2, PDF-X3, maximální spolehlivost při DTP a Prepress workflow (povinnost mít vložené fonty, . . . ).
- PDF-A verze pro archivaci skenovaných předloh.
- PDF-ISO (upravená verze 1.7 schválená institucí ISO)
- FAQ[102].

[101] http://www.pdf-x.com/
[102] http://www.pdf-x.com/pdfs/pdfx-faq.pdf



## Osvit



- Vytvoření předlohy na filmu (trvanlivost, odolnost vůči dilataci).
- Při barevném tisku v barevném modelu CMYK tedy 4 pláty pro každou stranu, plus případně direktní barvy (Pantone vzorník).
- Parametry osvitu jsou dány možnostmi konkrétní osvitové jednotky a jeho RIPu: typy bodů, počet linek na palec – lines per inch (LPI), nastavení rastrů, otočení.
- Autotypická síť bodů, autotypický rastr, frekvenčně modulovaný rastr.
- Běžné parametry otočení barev: nejtmavší blacK: 45°, Cyan: 15°, Magenta: 75°, Yellow: 0°.
- *Moiré*: pravidelné rušivé vzory při chybách otočení.
- Snímkové a pásové RIPy.



## Předtisková příprava a tisk: možný postup



- Z. Dvořáková: DTP a předtisková příprava, CPress 2008.
- Zpracování grafiky, obrázky, konverze a separace barev.
- Vytvoření elektronických předloh.
- Elektronická montáž a osvit.
- Vytvoření tiskové desky
- Ofsetový tisk.
- Dokončovací zpracování: skládání, ořez, vazba, lepení.

## Vazba, lepení, distribuce



- ☞ Vazba šitá, lepená.
- ☞ V1 na stříšku, do hřbetu.
- ☞ V2 lepená.
- ☞ V3 bloková brožura, šitá drátem naplocho.
- ☞ Distribuce, velkoobchod, maloobchod ⟶ *čtenář*.



## Archová montáž



- ☞ Montáž stránek archu obsahující soutiskové a ořezové značky potřebné pro vlastní tisk, skládání, ořez a falcování.
- ☞ Ruční (dělá se obvykle až v tiskárně pro konkrétní tiskový stroj).
- ☞ Elektronická: nad PDF či PS soubory před osvitem nebo přímo v aplikaci generující elektronickou verzi předlohy (dvips, InDesign).



## Paralelní publikování na WWW a jinak



- ☞ Důležitost specifického designu pro různá výstupní zařízení.
- ☞ Konverze z LATEXu do HTML či MATHML: LATEX2html, TEX4ht (Eitan Gurari), modul do Apache (Vildová).
- ☞ Konverze obrázků: `imagemagick`, `ghostscript`, `bm2eps`, `jpeg2ps`.
- ☞ Tisk html dokumentů (CSS), `html2ps`, FO procesory (via ConText apod.).
- ☞ Přímá sazba XML/(HTML) dokumentů (Context).
- ☞ Sazba databázových výstupů (elektronické pošty, faktur, vysvědčení, novinových článků).
- ☞ Problémy konverzí. Nutnost archivace zdrojů maximální kvality (CD Všech pět pohromadě, DVD 10@FI).



## Tiskové techniky



- ☞ Tisk z výšky, flexotisk.
- ☞ Tisk z hloubky.
- ☞ Tisk z plochy, litografie, ofset, světlotisk. Předlohou pauzák či film otočený zrcadlově.
- ☞ Sítotisk.
- ☞ Computer to Plate – přímý výstup tiskové desky (bez nutnosti svítit film).
- ☞ Digitální tisk, Print on Demand – tisk na požádání: hospodárný tisk malých až středních nákladů (do 1000 ks). Data jsou digitálně zpracovávána až po tisk, bez filmů a tiskových desek.

# Závěr

“We are all apprentices in a craft
where no-one ever becomes a master.”
*Ernest Hemingway*

☞ Úcta k tradicím.

☞ Osvěta, praxe.

“The road to wisdom?
Well it’s plain and simple to express:
Err and err and err again,
but less and less and less.”
*Piet Hein*

 © Petr Sojka, 2000–2008