

TVORBA ELEKTRONICKÝCH KNIH SYSTÉMEM TEX4EBOOK

Michal Hoftich¹

Přednáška pro CSTUG 2015

¹<michal.h21@gmail.com>

Ústřední knihovna PedF UK

Formáty elektronických knih

System tex4ht

Úvod do tex4ebook

Příklady

FORMÁTY ELEKTRONICKÝCH KNIH

- ePUB
 - nejrozšířenější podpora ve čtečkách
 - některé nepodporují ani kaskádové styly
- mobi
 - podpora ve čtečkách Kindle od Amazonu
 - vzniká konverzí z ePUBu programem kindlegen
- ePUB3

- Nové standardy
 - html5
 - mathml
 - CSS 3
 - SMIL
 - SVG

- fixed layout
- javascript
 - interaktivní prvky
 - vylepšení dokumentu
- struktura dokumentu
 - (frontmatter, mainmatter, backmatter)
 - dělení dokumentu (volume, part, chapter, subchapter)
- logické značkování
 - rejstříky
 - slovníky
- profily
 - edupub

<https://github.com/michal-h21/calzone>

The company is testing enhancements to private browsing in Firefox designed to block website elements **that could be used by third parties to track browsing behavior across sites. Most major browsers, Firefox included, have a Do Not Track option, though many companies do not honor it.**

Průliš *žluťoučký kůň* Mozillas experimental tool is designed to block outside parties like ad networks or analytics companies from **tracking users** through cookies and browser fingerprinting.

Its available in the Firefox Developer Edition on Windows, Mac and Linux, and Firefox Aurora on Android, Mozilla said Friday.

Mozillas experimental tool is designed to block outside parties like ad networks or analytics companies from tracking users through cookies and browser fingerprinting.

Its available in the Firefox Developer Edition on Windows, Mac and Linux, and Firefox Aurora on Android, Mozilla said Friday.

The tool is in prebeta, although it might be

- Nástroj Aeneas umí sesynchronizovat čtenou knihu s textem
- <http://www.readbeyond.it/aeneas/>
- Čtečka Menestrello zvýrazňuje právě čtenou pasáž v textu



<https://github.com/michal-h21/calzone>

The company is testing enhancements to private browsing in Firefox designed to block website elements **that could be used by third parties to track browsing behavior across sites. Most major browsers, Firefox included, have a Do Not Track option, though many companies do not honor it.**

Příliš **žluťoučký** kůň Mozillas experimental tool is designed to block outside parties like ad networks or analytics companies from *tracking users* through cookies and browser fingerprinting.

Its available in the Firefox Developer Edition on Windows, Mac and Linux, and Firefox Aurora on Android, Mozilla said Friday.

Mozillas experimental tool is designed to block outside parties like ad networks or analytics companies from tracking users through cookies and browser fingerprinting.

Its available in the Firefox Developer Edition on Windows, Mac and Linux, and Firefox Aurora on Android, Mozilla said Friday.

The tool is in prebeta, although it might be incorporated into future versions of Firefoxs main browser.

Pozor!

Platí zásada, že dokument musí být použitelný i ve čtečce bez podpory JS

- čtečky s e-inkem ho prakticky nepodporují
- iBooks
- Adobe Digital Editions
- Android - Gitden
- Azardi
- Radium SDK
- Menestrello

- pro validaci existuje nástroj ePUBcheck
- <https://github.com/IDPF/epubcheck>

Trochu inline matematiky: $\sqrt{a^2 + b^2}$. Tady pokračuje další text.

$$\sin A \cos B = \frac{1}{2} [\sin(A - B) + \sin(A + B)]$$

$$\sin A \sin B = \frac{1}{2} [\sin(A - B) - \cos(A + B)]$$

$$\cos A \cos B = \frac{1}{2} [\cos(A - B) + \cos(A + B)]$$

Obrázek 3: Matematika ve formátu PNG. Azardi

Trochu inline matematiky: $\sqrt{a^2 + b^2}$. Tady pokračuje další text.

$$\sin A \cos B = \frac{1}{2} [\sin(A - B) + \sin(A + B)]$$

$$\sin A \sin B = \frac{1}{2} [\sin(A - B) - \cos(A + B)]$$

$$\cos A \cos B = \frac{1}{2} [\cos(A - B) + \cos(A + B)]$$

Obrázek 4: Matematika ve formátu PNG. Kindle

Trochu inline matematiky: $\sqrt{a^2 + b^2}$. Tady pokračuje další text.

$$\sin A \cos B = \frac{1}{2}[\sin(A - B) + \sin(A + B)]$$

$$\sin A \sin B = \frac{1}{2}[\sin(A - B) - \cos(A + B)]$$

$$\cos A \cos B = \frac{1}{2}[\cos(A - B) + \cos(A + B)]$$

Obrázek 5: Matematika ve formátu SVG. Azardi

Trochu inline matematiky:

$$\sqrt{a^2 + b^2}$$

. Tady pokračuje další text.

$$\sin A \cos B = \frac{1}{2} [\sin(A - B) + \sin(A + B)]$$
$$\sin A \sin B = \frac{1}{2} [\sin(A - B) - \sin(A + B)]$$
$$\cos A \cos B = \frac{1}{2} [\cos(A - B) + \cos(A + B)]$$

Obrázek 6: Matematika ve formátu SVG. Kindle

Trochu inline matematiky: $\sqrt{a^2 + b^2}$. Tady pokračuje další text.

$$\sin A \cos B = \frac{1}{2} [\sin (A - B) + \sin (A + B)]$$

$$\sin A \sin B = \frac{1}{2} [\sin (A - B) - \cos (A + B)]$$

$$\cos A \cos B = \frac{1}{2} [\cos (A - B) + \cos (A + B)]$$

Obrázek 7: Matematika ve formátu MathML. ADE

SYSTÉM TEX4HT

- <https://www.tug.org/tex4ht/>
- původní autor Eitan Gurari (1947–2009)
- systém vzniká od poloviny 90. let
- po smrti původního autora se o vývoj starají Karl Berry, CV Radhakrishnan a Michal Hoftich

- Pandoc
- LaTeXML
- LaTeX2HTML

- systém se skládá z množství kompilačních skriptů, které se ale liší jen v přednastavených hodnotách
- základní skript je `htlatex`
- například pro konverzi do ODT slouží `mk4ht oolatex`
- kompilace sestává ze tří základních kroků:

- pro podporované balíčky jsou nahrané **.4ht** soubory, které vkládají konfigurovatelné háčky na vhodná místa
- po vložení háčků se nahrají jejich konfigurace v závislosti na výstupním formátu

- zápis výstupních souborů
- konverze znakových sad
 - poměrně komplikovaný proces, potřebujeme doplňkové soubory pro fonty obsahující unicode entity nebo ASCII znaky pro jednotlivé znaky písma
- zachovává základní styly písem, podporuje jakákoli makra měnící vzhled písma
- příprava .lg souboru
- zápis .idv souboru obsahující stránky pro konverzi na obrázky

- výstup do CSS souboru
- konverze vložených obrázků z .dvi do výstupního formátu
- zpracování výstupních souborů externími programy (`xslt`, `tidy`, `xmllint`, `xtpipes`)
- kopírování souborů do cílového adresáře

- parametry pro každý jednotlivý krok se předávají kompilačnímu skriptu v závorkách
- základní forma pro dokument ve formátu utf8

Příklad

```
htlatex jmenosouboru.tex "xhtml,charset=utf-8"  
" -cmozhtf -utf8"
```

- soubor `tex4ht.sty` se nahrává ještě před načtením dokumentu, jeho nahrávání si dále řídí sám
- po zpracování preamble a nahrání všech balíčků se spouští `.4ht` soubory s vkládáním háčků pro dané balíčky, pokud existují

- soubor tex4ht.sty se nahrává ještě před načtením dokumentu, jeho nahrávání si dále řídí sám
- po zpracování preamble a nahrání všech balíčků se spouští .4ht soubory s vkládáním háčků pro dané balíčky, pokud existují

Pozor!

Protože se příkazy redefinují až na začátku dokumentu, příkazy volané v preamble dokumentu nejsou ještě redefinované

- konfigurace je důležitá hlavně pro příkazy se složitější strukturou nebo logické bloky (nadpisy, tabulky, seznamy apod.)
- pokud makra staví na základních prvcích, pro které už existuje podpora, nemusí být třeba přidávat pro ně jejich vlastní podporu

- konfigurace probíhá příkazem `\Configure`
- po vložení háčků se nahrají jejich konfigurace v závislosti na výstupním formátu
- další konfigurace je možné vložit do `.cfg` souboru

ÚVOD DO TEX4EBOOK

- <https://github.com/michal-h21/tex4ebook>
- staví na tex4ht a přidává:
- podporu pro ePUBová metadata
 - obsah
 - obálka
 - OPF soubor
- pro každou kapitolu nebo sekci je vytvořený samostatný soubor
- podporu build souborů pro make4ht

- <https://github.com/michal-h21/make4ht>
- build systém pro tex4ht, který řeší základní problémy:
- složité předávání parametrů pro htlatex a ostatní skripty
- pevně nastavené pořadí volání jednotlivých kroků kompilace
 - TeX je vždy volán třikrát
- podpora pro nástroje jako je **bibtex**, **xindy** a podobně
- snadná změna parametrů konverze obrázků
- funkční kopírování souborů do výstupního adresáře
- zpracování výstupních souborů filtrovacími funkcemi v jazyce Lua, nebo externími programy

PŘÍKLADY

```
\documentclass{article}
\usepackage[T1]{fontenc}
\usepackage[utf8]{inputenc}
\usepackage[czech]{babel}
\title{Základní dokument v češtině}
\author{Michal Hoftich}
\begin{document}
\maketitle
```

Příliš žluťoučký kůň úpěl ďábelské ódy

```
\end{document}
```

Základní postup

```
tex4ebook basic
```

Formát ePUB 3

```
tex4ebook -f epub3 basic
```

Zrychlená kompilace

```
tex4ebook -m draft basic
```

```

<div class="maketitle">
<h2 class="titleHead">Základní dokument v češtině
</h2><div class="author" ><span
class="ecrm-1200">Michal Hoftich</span></div><b>
<div class="date" ><span
class="ecrm-1200">1.</span><span
class="ecrm-1200"> </span><span
class="ecrm-1200">července 2015</span></div>
</div>
<!--l. 10--><p class="indent" >
Příliš žluťoučký kůň úpěl ďábelské ódy </p>

```

```
\documentclass{article}
\usepackage[T1]{fontenc}
\usepackage[utf8]{inputenc}
\usepackage[czech]{babel}
\begin{document}
{\huge \texttt{tex4ht} umí zachovat
  \small základní {\bfseries formát}
  písem}
\end{document}
```



tex4ht umí zachovat základní **formát** písem

Obrázek 8: Velikost a řez písma jsou zachovány

```
\documentclass{article}  
\usepackage{fontspec}  
%\usepackage[czech]{babel}  
\title{Základní dokument v češtině}  
\author{Michal Hoftich}  
\begin{document}  
\maketitle
```

Příliš žluťoučký kůň úpěl ďábelské ódy

```
\end{document}
```

Kompilace

```
tex4ebook -m draft -l fontspec.tex  
"new-accents"
```

A její výsledek

```
--- error --- Can't find/open file  
"file:lmroman12-regular:script=latn;  
+trep;+tlig;".tfm'  
Make4ht: Fatal error. Command tex4ht  
returned exit code 256
```


- je třeba zamezit, aby se vůbec načetly
- tři možná řešení
 1. použít rozdílné šablony pro preamble dokumentu, samotný text vkládat pomocí `\include`
 2. použít podmínku detekující tex4ht
 3. alternativní loader balíčků

```
\documentclass{article}  
\ifdefined\HCode  
\usepackage[T1]{fontenc}  
\usepackage[utf8]{luainputenc}  
\else  
\usepackage{fontspec}  
\fi  
\usepackage[czech]{babel}  
\title{Základní dokument v češtině}  
...
```

```
\documentclass{article}  
\usepackage{alternative4ht}  
\altusepackage{fontspec}  
  \setmainfont{TeX Gyre Termes}  
  \newfontfamily\greekfont{Linux Libertine O}  
  ...
```

```
\documentclass{article}
...
\usepackage{biblatex}
\addbibresource{xampl.bib}
...
\tableofcontents
\section{Úvod}
Příliš žluťoučký kůň \textit{úpěl}
\textbf{dábelské ódy} \parencite{article-full}
\section{Další kapitola}
Nějaká matematika  $\sqrt{a^2 + b^2}$ 
\section{Literatura}
\printbibliography
```

Chybné převedení diakritiky

```
<p class="noindent" >Příliš žluťoučký kůň <span  
class="ecti-1000">úp</span><span  
class="ecti-1000">ěl </span><span  
class="ecbx-1000">d'</span><span  
class="ecbx-1000">ábel'sk</span><span  
class="ecbx-1000">é </span><span  
class="ecbx-1000">ódy </span>
```

- umožňují vkládání konfigurací pro háčky a CSS instrukce
- základní struktura

```
% zde můžeme vkládat balíčky
\Preamble{xhtml,volby pro tex4ht.sty}

...
\begin{document}

...
\EndPreamble
```

```
\Preamble{xhtml}  
\Configure{textbf}{\NoFonts\HCode{<strong>}}  
{\HCode{</strong>}\EndNoFonts}  
\Configure{textit}{\NoFonts\HCode{<em>}}  
{\HCode{</em>}\EndNoFonts}  
\begin{document}  
\EndPreamble
```

```
<p class="noindent" >Příliš žluťoučký kůň  
<em>úpěl</em> <strong>dábelské ódy</strong>
```



```
\Preamble{xhtml}  
\Configure{textbf}{\NoFonts\HCode{<strong>}}  
{\HCode{</strong>}\EndNoFonts}  
\Configure{textit}{\NoFonts\HCode{<em>}}  
{\HCode{</em>}\EndNoFonts}  
\Css{h3{color:blue;}}  
\begin{document}  
\EndPreamble
```

1 Úvod

Příliš žluťoučký kůň *úpěl* **dábelské** *ódy* [1]

Obrázek 9: Modrý nadpis

Zkusme přidat kompletní styl pro responzivní design a fonty

- responzivní styl Scale.css
(<https://github.com/viljamis/Scale>)
- font EB Garamond

Pro usnadnění práce si vytvoříme pomocné balíčky

```
\newcommand\AddFile[1]{\special{t4ht+@File: #1}}%  
\newcommand\AddCss[1]{%  
\AddFile{#1}%  
\Configure{@HEAD}{%  
\HCode{<link rel="stylesheet" type="text/css"  
href="#1" />}}  
}
```

```
\RequirePackage{include4ht}
\newcommand\AddFontFace[4]{%
\Css{@font-face {
  font-family: #1;
  src: local("#2"),
  url('#3');
  #4
}}
\AddFile{#3}}
\edef\CurrentFontFamily{rmfamily}
\newcommand\SetFontFamily[1]{%
\edef\CurrentFontFamily{#1}}
```

```
\newcommand\NormalFont[2]{  
  \AddFontFace{\CurrentFontFamily}{#1}{#2}  
  {font-weight: normal;font-style: normal;}}  
\newcommand\BoldFont[2]{  
  \AddFontFace{\CurrentFontFamily}{#1}{#2}  
  {font-weight: bold;font-style: normal;}}  
\newcommand\ItalicFont[2]{  
  \AddFontFace{\CurrentFontFamily}{#1}{#2}  
  {font-weight: normal;font-style: italic;}}
```

```
\RequirePackage{include4ht}
\RequirePackage{addfont4ht}
\Preamble{xhtml}
\AddCss{scale.css}
...
\NormalFont{EBGaramond}{EBGaramond12-Regular.woff}
\BoldFont{EBGaramond}{EBGaramond12-Italic.woff}
\ItalicFont{EBGaramond}{EBGaramond12-Italic.woff}
\Configure{@HEAD}{\HCode{<style type='text/css' >
    \Hnewline body{font-family:rmfamily,
        "EBGaramond", sans-serif;}\Hnewline
    </style>}}
\EndPreamble
```

```
tex4ebook -f epub3 -c epub3.cfg -m draft basic.tex
```


I Úvod

Příliš žlutoučký kůň *úpěl ďábelské ódy* [I]

Obrázek 10:

```
Make:add("biber","biber ${input}")  
Make:htlatex {}  
Make:biber {}  
Make:htlatex {}  
Make:image("png$",  
"dvipng -bg Transparent -T tight -o ${output}"..  
"-pp ${page} ${source}")
```

```
tex4ebook -f epub3 -c epub3.cfg -m draft  
-e pokus.mk4 basic.tex
```

2 Další kapitola

Nějaká matematika $\sqrt{a^2 + b^2}$

Obrázek 11: Dalo by se hrát s konfiguračními volbami, ale lepší je použít mathml

Pozor!

Scale.css upravuje nastavení obrázků a úplně rozhodí inline matematiku. Proto v ukázce není použit

```
tex4ebook -c epub.cfg -f epub3 -m draft  
-e pokus.mk4 basic mathml
```

2 Další kapitola

Nějaká matematika $\sqrt{a^2 + b^2}$

- Máme dvě možnosti:
 1. použít příkaz `\coverimage`
 - to vyžaduje explicitní vložení balíčku `tex4ebook` do dokumentu
 - obálka se zobrazí v dokumentu
 2. použít příkaz `\CoverMetadata` v konfiguračním souboru
 - výhodou je, že nemusíte upravovat dokument
 - obálka se zobrazí pouze ve čtecí aplikaci

- výstupní soubory můžeme zpracovat externími příkazy, nebo funkcemi v jazyce Lua
- efektivní zpracování umožňují filtry


```
\documentclass{article}
...
\begin{document}
Co dělat v případě, že se
{\bfseries nám nelíbí výstup?}
\hrule
```

Použijeme filtry!

```
\end{document}
```

```
<p class="noindent">Co dělat v případě, že se  
<span class="ecbx-1000">n</span><span  
class="ecbx-1000">ám nel</span><span  
class="ecbx-1000">íb</span><span  
class="ecbx-1000">í v</span><span  
class="ecbx-1000">ýstup?</span>_____</p>  
<p class="indent">    Použijeme filtry!
```

```
local filter = require "make4ht-filter"  
local process = filter{"cleanspan", "hruletohr"}  
Make:htlatex()  
Make:htlatex()  
Make:match("html$",process)  
Make:match("html$",  
"tidy -m -utf8 -asxhtml -q -i ${filename}")
```

```
<body>
<p class="noindent">Co dělat v případě, že se
<span class="ecbx-1000">nám nelíbí výstup?
</span>
</p>
<hr class="hrule" />

<p class="indent">Použijeme filtry!</p>
</body>
```

- profil pro ePub 3 určený pro vzdělávací materiály
- <http://www.idpf.org/epub/profiles/edu/spec/>
- přidává sémantická metadata

- vytvoříme balíček pro vkládání učitelských poznámek
- poznámky se vypíší pouze učitelům

- balíček bude mít volitelný argument **teacher**
- učitelské poznámky se vypíší pouze pokud je balíček vložen s tímto argumentem
- dva řídící soubory obsahující kompletní preamble, vkládají TeXový soubor, který je pro učitele i studenty stejný

```
\documentclass{article}  
...  
\usepackage[teacher]{edupub}  
\begin{document}  
\include{text}  
\end{document}
```



```
\documentclass{article}  
...  
\usepackage[]{edupub}  
\begin{document}  
\include{text}  
\end{document}
```

```
\newcommand\odpoved[1]{\\ \teacherinfo{#1}}  
\begin{enumerate}  
\item Jak se jmenuje náš největší had?  
\odpoved{Užovka stromová}  
\end{enumerate}
```

```
\ProvidesPackage{edupub}
\RequirePackage{kvoptions}
\RequirePackage{etoolbox}
\newbool{teacher}
\boolfalse{teacher}
\newcommand\teacherinfo[1]{}
\DeclareVoidOption{teacher}{%
  \renewcommand\teacherinfo[1]{%
    \edupub@print@teacherinfo{##1}}
  \booltrue{teacher}}
\newcommand\edupub@print@teacherinfo[1]{#1}
\ProcessKeyvalOptions*
```

```
\ifbool{teacher}{%  
  \Configure{OpfMetadata}  
  {\HCode{<dc:type>teacher-edition</dc:type>}}  
}{%}
```

```
\NewConfigure{teacherinfo}{2}

\let\old:teacherinfo\edupub@print@teacherinfo

\renewcommand\edupub@print@teacherinfo[1]{
  \a:teacherinfo
  \old:teacherinfo{#1}
  \b:teacherinfo
}
\Configure{teacherinfo}
{\HCode{<span epub:type="answer">}}
{\HCode{</span>}}
```

student

```
<li class="enumerate" id="x1-3x1">  
Jak se jmenuje náš největší had?<br  
class="newline" /></li>
```

učitel

```
<li class="enumerate" id="x1-3x1">  
Jak se jmenuje náš největší had?<br  
class="newline" />  
<span epub:type="answer">Užovka stromová </span>  
</li>
```

Děkuji za pozornost