

# Single-source publishing in multiple formats for different output devices

Petr Sojka, Michal Růžicka

Faculty of Informatics  
Masaryk University, Brno

Talk presented at EuroBachTeX in Bachotek on May 1st, 2007

- 1 Single-source publishing
  - Content and Form
  - Mark up
- 2 PDF versions
  - PDF for printing
  - PDF for on-screen viewing
- 3 Versions for the web delivery
  - XHTML + MATHML
- 4 On-the-fly document generation
  - ‘Yellow book’ of FI MU
  - Math digital library with JBIG2
  - Summary







Značkování  
Markup

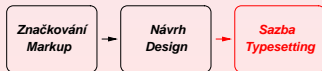
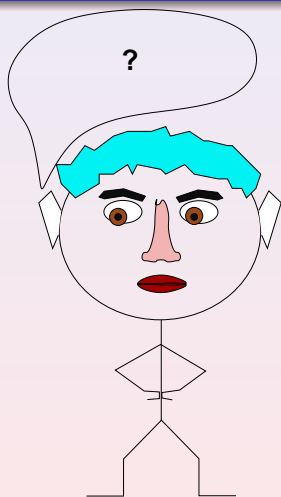


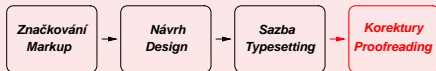


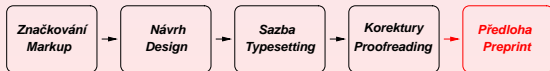
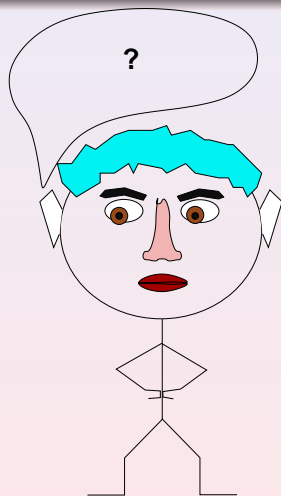
Značkování  
Markup

Návrh  
Design



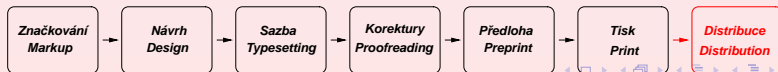
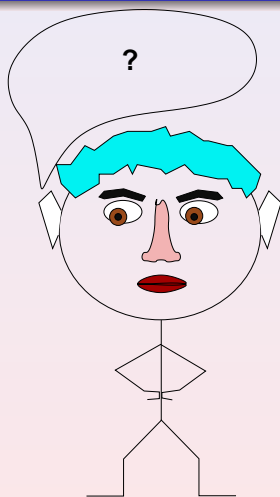


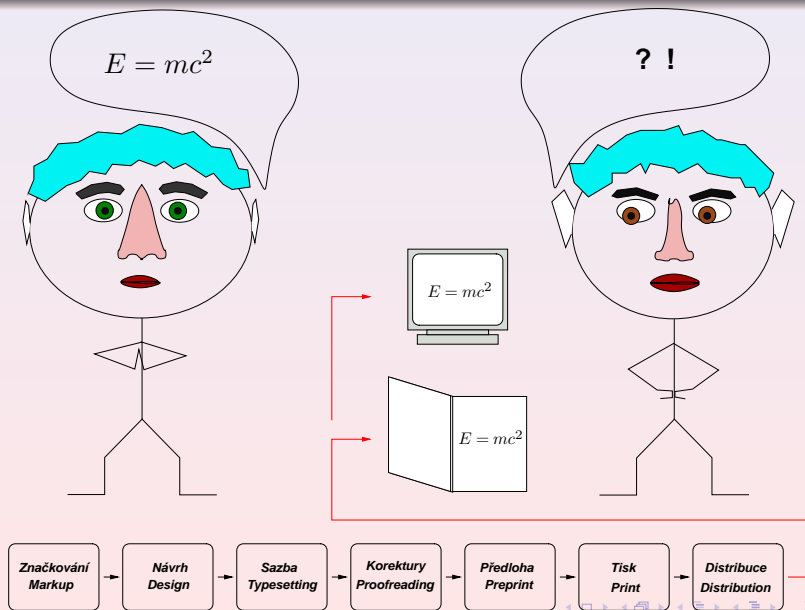












# Content and Form

Discover the outer logic of the typography  
in the inner logic of the text.  
— Robert Bringhurst

- Document = content + form.



# Content and Form

Discover the outer logic of the typography  
in the inner logic of the text.  
— Robert Bringhurst

- Document = content + form.
- Content should be marked up in author's preferences of domain language.



# Content and Form

Discover the outer logic of the typography  
in the inner logic of the text.  
— Robert Bringhurst

- Document = *content* + form.
- Content should be marked up in author's preferences of domain language.
- Form (appearance) should reflect the design, it should use the graphical means consistently (sameness).



# Content and Form

Discover the outer logic of the typography  
in the inner logic of the text.  
— Robert Bringhurst

- Document = content + form.
- Content should be marked up in author's preferences of domain language.
- Form (appearance) should reflect the design, it should use the graphical means consistently (sameness).
- Possibilities of a form of a document are constrained by an output devices (paper, LCD monitor, PDA).



# Content and Form

Discover the outer logic of the typography  
in the inner logic of the text.  
— Robert Bringhurst

- Document = content + form.
- Content should be marked up in author's preferences of domain language.
- Form (appearance) should reflect the design, it should use the graphical means consistently (sameness).
- Possibilities of a form of a document are constrained by an output devices (paper, LCD monitor, PDA).
- Single source publishing allows structured aggregation of content and form markup and cost-effective maintenance.



## Mark up (author)

```
&\elevenit I\kern.7ptillustrations by\cr  
&DU\kern-1ptANE BIBBY\cr  
\noalign{\vfill}  
&\setbox0=\hbox{\manual77}%  
\setbox2=\hbox to\wd0{\hss\manual6\hss}%  
\raise2.3mm\box2\kern-\wd0\box0\cr % A-W logo  
&ADDISON\kern.1em--WESLEY\cr  
&PUBLISHING COMP\kern-.13emANY\kern-1.5mm\cr
```

?



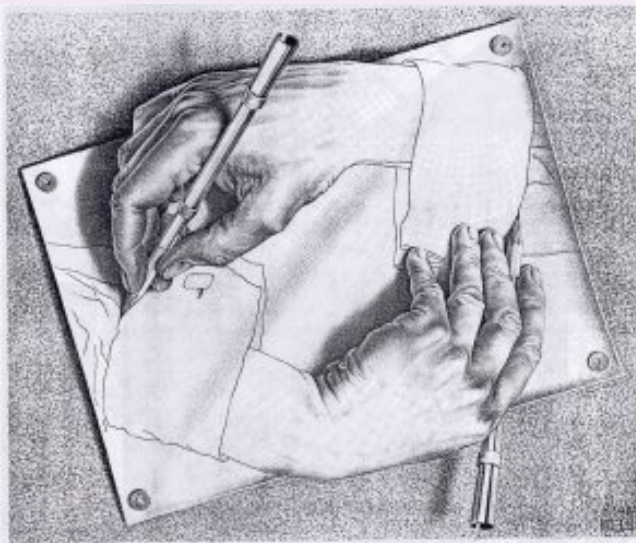
## Mark up (author)

```
&\elevenit I\kern.7ptllustrations by\cr  
&DU\kern-1ptANE BIBBY\cr  
\noalign{\vfill}  
&\setbox0=\hbox{\manual77}%  
\setbox2=\hbox to\wd0{\hss\manual6\hss}%  
\raise2.3mm\box2\kern-\wd0\box0\cr % A-W logo  
&ADDISON\kern.1em--WESLEY\cr  
&PUBLISHING COMP\kern-.13emANY\kern-1.5mm\cr
```

?

NO! (for single source publishing for multiple outputs)

# Mark up (author)



## Mark up (cont.)

Data cannot be used at a finer grain than it is marked up at. — Rick Jelliffe

- Habit of mixing visual and logical markup by  $\text{\TeX}$  authors (as opposed to XML authoring).



## Mark up (cont.)

Data cannot be used at a finer grain than it is marked up at. — Rick Jelliffe

- Habit of mixing visual and logical markup by  $\text{\TeX}$  authors (as opposed to XML authoring).
- XML (MathML) is a ‘no way’ for mathematical authors for content authoring.



## Mark up (cont.)

Data cannot be used at a finer grain than it is marked up at. — Rick Jelliffe

- Habit of mixing visual and logical markup by  $\text{\TeX}$  authors (as opposed to XML authoring).
- XML (MathML) is a ‘no way’ for mathematical authors for content authoring.
- Any fixed markup dictionary is not sufficient to model and express the content domain.



## Mark up (cont.)

Data cannot be used at a finer grain than it is marked up at. — Rick Jelliffe

- Habit of mixing visual and logical markup by  $\text{\TeX}$  authors (as opposed to XML authoring).
- XML (MathML) is a ‘no way’ for mathematical authors for content authoring.
- Any fixed markup dictionary is not sufficient to model and express the content domain.
- Significant cleanup needed for ‘print only’ authored sources.



Make all visual distinctions as subtle as possible,  
but still clear and effective.  
— Edward R. Tufte

For every output version, design parameters and macros are, as usual,  
written in separate conditional branch. Simple source code example:

```
\newif\ifprint
\printfalse % Non-print version.
%\printtrue % Print version.

\ifprint
  \hypersetup{colorlinks=false,
              pdfborder={0 0 0}}
  % Center mirrored document pages on A4
  % paper and add crop marks.
  \usepackage[cam,a4,center,mirror]{crop}
\fi
```





# PDF for on-screen viewing

Compared to typical laser printer output device:

- different aspect ratio



# PDF for on-screen viewing

Compared to typical laser printer output device:

- different aspect ratio
- colors



# PDF for on-screen viewing

Compared to typical laser printer output device:

- different aspect ratio
- colors
- order of magnitude lower resolution



## PDF for on-screen viewing

Compared to typical laser printer output device:

- different aspect ratio
- colors
- order of magnitude lower resolution
- hyperlinks



## PDF for on-screen viewing

Compared to typical laser printer output device:

- different aspect ratio
- colors
- order of magnitude lower resolution
- hyperlinks
- navigation toolbars,...

Same line breaking possible in some designs for lazy typesetters.



## Version for the web (browser) I: HTML

- Many tools available: TeX2page, Tralics, TeX4ht, LaTeXXML, or LaTeX2HTML.



## Version for the web (browser) I: HTML

- Many tools available: TeX2page, Tralics, TeX4ht, LaTeXXML, or LaTeX2HTML.
- Pros and cons

# Version for the web (browser) I: HTML

- Many tools available: TeX2page, Tralics, TeX4ht, LaTeXXML, or LaTeX2HTML.
- Pros and cons
- TeX4ht

```
htllatex filename.tex 'html'
```





# Version for the web (browser) I: HTML

- Many tools available: TeX2page, Tralics, TeX4ht, LaTeXXML, or LaTeX2HTML.
- Pros and cons
- TeX4ht

```
htllatex filename.tex 'html'
```

Example in browser (math as bitmaps)



## Version for the web (browser) I: CSS

```
\ifweb % In case of web format output
  \Css{% CSS code definition block
    .theorem {
      background-color: \#FFFFFF;
      border: 1px solid;
      border-color: \#0000FF;
    }
  }
% In the resulting HTML document mark
% "<div class="theorem">" is placed
% before each "theorem" environment
% and mark "</div>" after that.
\ConfigureEnv{theorem}
  {\HCode{<div class="theorem">}}
  {\HCode{</div>}}
  {}{}
\fi
```

## Version for the web (browser) I: CSS

```
\ifweb % In case of web format output
  \Css{% CSS code definition block
    .theorem {
      background-color: \#FFFFFF;
      border: 1px solid;
      border-color: \#0000FF;
    }
  }
```

```
% In the resulting HTML document mark
% "<div class="theorem">" is placed
% before each "theorem" environment
% and mark "</div>" after that.
```

```
\ConfigureEnv{theorem}
  {\HCode{<div class="theorem">}}
  {\HCode{</div>}}
  {}{}
\fi
```

```
% In document, same LaTeX markup
% is used for all versions:
\begin{theorem}
  Function~ $f$  have only one limit
  in point~ $[x_{0},y_{0}]$ .
\end{theorem}
```

## Version for the web (browser) II: XHTML+MathML

```
htlax filename.tex 'xhtml,mozilla'
```



## Version for the web (browser) II: XHTML+MathML

```
htllatex filename.tex 'xhtml,mozilla'
```

- *scalable*



## Version for the web (browser) II: XHTML+MathML

```
htllatex filename.tex 'xhtml,mozilla'
```

- *scalable, searchable (math as text, not pictures)*



## Version for the web (browser) II: XHTML+MathML

```
htllatex filename.tex 'xhtml,mozilla'
```

- *scalable, searchable* (math as text, not pictures)

```
$M=\{x|x \mbox{is odd}\}$
```

*has to be used instead of correct  $\TeX$  code*

```
$M=\{x|x$ is odd $\}$
```



## 'Yellow book' of FI MU example

- content partially generated from database of IS MU





## 'Yellow book' of FI MU example

- content partially generated from database of IS MU
- some content acquired by validating web forms (course syllabæ,...)



## 'Yellow book' of FI MU example

- content partially generated from database of IS MU
- some content acquired by validating web forms (course syllabæ,...)
- print and screen format generated from  $\text{\LaTeX}$  generated from XML via XSL.



# Math digital library with JBIG2 example

- 250.000 pages to digitize, scanned as 600 DPI TIFF



## Math digital library with JBIG2 example

- 250.000 pages to digitize, scanned as 600 DPI TIFF
- OCR (FineReader) to have text under image layer for searching



## Math digital library with JBIG2 example

- 250.000 pages to digitize, scanned as 600 DPI TIFF
- OCR (FineReader) to have text under image layer for searching
- OCR (InftyReader) to have MathML or  $\LaTeX$  under image layer for searching



## Math digital library with JBIG2 example

- 250.000 pages to digitize, scanned as 600 DPI TIFF
- OCR (FineReader) to have text under image layer for searching
- OCR (InftyReader) to have MathML or  $\LaTeX$  under image layer for searching
- image recompressing with JBIG2ENC in JBIG2 format to save space and bandwidth (under preparation, up to 80% size reduction)



# Summary



# Summary



Značkování  
Markup





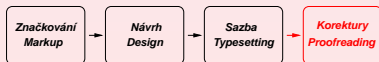
# Summary



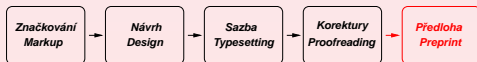
# Summary



# Summary



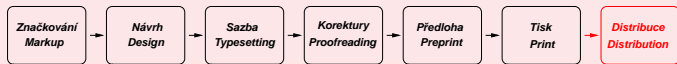
# Summary



# Summary



# Summary



# Summary

