L$_A$T$_E$X and MS Word Documents and Their Presentation in Web Environment

MILOŠ ŠEDA*, JINDŘIŠKA ŠEDOVÁ**
*Institute of Automation and Computer Science
Brno University of Technology, Faculty of Mechanical Engineering
Technická 2, 616 69 Brno
CZECH REPUBLIC
seda@fme.vutbr.cz
**Department of Law
Masaryk University, Faculty of Economics and Administration
Lipová 41a, 602 00 Brno
CZECH REPUBLIC
jsedova@econ.muni.cz

Abstract: - In distance learning, and, of course, not only in it, web tools play gradually increasing role. A frequent problem is to use traditional documents, mainly prepared in Microsoft Word or L$_A$T$_E$X, and transform them, if possible, by a simple way to be included in web pages. This paper discusses some freeware or shareware programs and tools available from the Internet and the ways they were used, besides sophisticated and more expensive solutions, for simple and fast preparation of teaching materials to be used for distance education and lifelong learning.

Key-Words: - markup languages, graphic formats, frames, script

1 Introduction

In modern teaching processes, most documents are available on the Internet and they are frequently included in the PHP-MySQL [8,9] environment. The teaching materials can also be prepared in an e-learning form where their readers pass tests and only a sufficient number of correct answers qualifies him or her to continue reading the following chapter. A well-known tool of e-learning programs is Moodle.

Besides PowerPoint presentations, Web browsers are also used by lecturers and thus it is necessary to have materials prepared in a www form. However, experienced teachers have many documents written in text processors such as Microsoft Word, mathematicians and physicians prefer a T$_E$X platform and face the problem of making their texts available.

Though Microsoft Word has its own tool to transform the .doc format to a www one, it took more than 100 text lines to convert (and generate into a .mht file) the text containing merely the headline of this paper. Most of them require complex style definitions. It is evident that this effect can be provided by simple HTML directives. Additionally, the www code generated by Microsoft Word is quite unsuitable for modifications. Therefore, there is a point in exploring other ways of transforming MS Word documents to www formats. We will extend our considerations also to T$_E$X documents.

2 Conversion of Text Documents

Many T$_E$X (L$_A$T$_E$X, AMST$_E$X, …) and HTML directives have similar form. Let us assume, e.g., an ordered list with two items in T$_E$X and directives written in bold style as follows:

\begin{enumerate}
\item Seda, M.: A Comparison of Roadmap and Cell Decomposition Methods in Robot Motion Planning. **{\it** WSEAS Transactions on Systems and Control**}**, Issue 2, Vol. 2, 2007, pp. 101-108.
\item Seda, M.: Fuzzy All-Pairs Shortest Paths Problem. In: Reusch, B. (ed.): **{\it** Computational Intelligence, Theory and Applications**}**. Springer-Verlag, Berlin (Germany), 2006, pp. 395-404.
\end{enumerate}

Its HTML equivalent has the following form:

<ol>

Seda, M.: A Comparison of Roadmap and Cell Decomposition Methods in Robot Motion Planning. <i>WSEAS Transactions on Systems and Control</i>, Issue 2, Vol. 2, 2007, pp. 101-108.
Seda, M.: Fuzzy All-Pairs Shortest Paths Problem. In: Reusch, B. (ed.): <i>Computational Intelligence, Theory and Applications</i>. Springer-Verlag, Berlin (Germany), 2006, pp. 395-404.

From the similar directives for ordered lists, their items and italics style in T_EX and HTML, it is evident that it is not too difficult to implement a conversion program for their mutual replacement. These analogies can be demonstrated in many other statements of these "markup" languages.

A conversion tool that allows documents written in L_AT_EX to become part of the World Wide Web was developed by Nikos Drakos from University of Leeds [3]. Its author called it LaTeX2HTML Translator. In addition to basic statements, such as formatting information with equivalent "tags" in HTML (lists, quotes, paragraph breaks, type styles, etc.), conversion of cross-references, citations, footnotes, tables of contents, and lists of figures and tables is provided as well.

Of course, there are also elements of T_EX languages that are not so easy to convert or their direct conversion is impossible. This can mainly be said about tables, mathematical formulas and pictures. The correspondence between L_AT_EX and HTML tables is not straightforward even for very simple tables with left aligned cells in two rows and two columns without specifications concerning the widths and heights of cells as shown in the following example:

1.L_AT_EX

```
\begin{table}[htpb]
 \begin{tabular}{|l|l|}
   \hline
    Graph Theory & 5, MS\\
   \hline
    Automation    & 4, MS\\
   \hline
 \end{tabular}
\end{table}
```

2.HTML

```
<table TABLE border="1">
 <tr>
   <td>Graph Theory</td>
   <td>5, MS</td>
 </tr>
 <tr>
```

```
   <td>Automation</td>
   <td>4, MS; 3, BS</td>
 </tr>
</table>
```

Let us consider the following formula:

$$\forall k \in M - \{1\} : v_{1k} = \sum_{r=1}^{k-1} \sum_{i=1}^{n} p_{ir}.x_{1i} \qquad (1)$$

Its L_AT_EX equivalent (including automatic numbering) has the following form:

```
\begin{equation}
    \forall k\in M-\{1\}: v_{1k} =
        \sum_{r=1}^{k-1} \sum_{i=1}^{n} p_{ir}x_{1i}
\end{equation}
```

Unfortunately, there are no HTML equation directives.

2.1 Transfer of Pictures

Regarding pictures, there are two possible ways of inserting them into L_AT_EX documents: The first one is based on directly drawing a picture in L_AT_EXCAD available in the L_AT_EX environment, while the second one on inserting images saved in graphical formats such as .JPEG, .GIF, .PNG etc.

L_AT_EXCAD is a very simple graphical tool, to a certain extent similar to the one implemented in Microsoft Word, that includes selecting the line thickness (thin, thick, THICK) and includes statements for drawing various types of lines (normal, dash, dotted, vector), boxes (frame box, shadow frame box, dash box, shaded box), curves (circle, shaded circle, oval, ellipse, Bezier curves, arcs), dots (dot, thick dot, hollow dot), braces (left brace, right brace, over brace, under brace), and text boxes (normal, framed, shadow box, double box, oval box, thick oval box). Unfortunately, L_AT_EXCAD statements do not have equivalents in HTML.

The second possible way of including pictures into MiKT$_EX$ [6] (it is L_AT_EX implemented for direct use in Windows) demonstrates the following source code outline. It provides the inclusion of a centred image saved in the graphic file Fig1.jpg with defined size and label situated under the picture. The directive \usepackage{graphicx} maps all tools necessary to implement this operation.

```
\usepackage{graphicx}
```

```
\begin{figure}[h]
  \centering
    \includegraphics*[bb=0 0 370 231]{Fig1.jpg}
```

```
    \caption{ESMT as a union of full Steiner trees}
      \label{fig:Fig1}
\end{figure}
```

In this case, figures can be very simply inserted into HTML. For Fig1.jpg from the previous example, for instance, the statements can be used as follows:

```
<center>
 <br>
 <img src="Fig1.jpg" width="370" height="231"
      border="0">
 <br>ESMT as a union of full Steiner trees
 <br><br>
</center>
```

Therefore, we must find another way of converting tables, mathematical formulas and pictures from L_AT_EXCAD to HTML. The simplest way seems to be to convert them to images which will be placed at the correct positions in the final HTML document. This approach is also used in LaTeX2HTML Translator [3].

The conversion to images can be very simply provided by IrfanView, which is a freeware graphic viewer for Windows developed by Irfan Skiljan from Bosnia and Herzegovina. Now, version 4.00 of this program is available [7]. It is sufficient to select a table, mathematical formula or picture to clipboard, insert its contents to IrfanView (provided that .CLP option is checked in the Options-Properties-Extensions window) and then save it in one of the graphical formats checked in the list of those available. We can choose from more than 60 graphical formats, e.g. .JPG/.JPEG, .GIF, .PNG, .TIF/.TIFF, .WMF, .BMP and .PCX. This approach is shown for formula (1) in the following three figures.
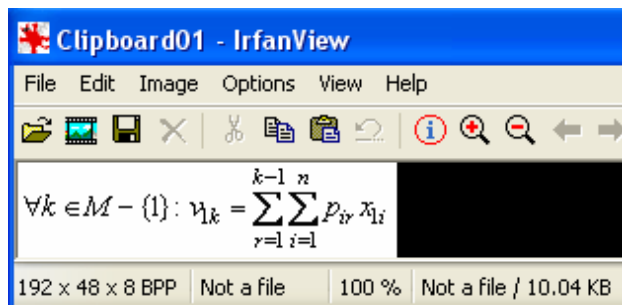
Fig. 1. Formula (1) in clipboard inserted in IrfanView

Let us note that this program has other useful functions such as resize. Photographers will appreciate its many options in setting the brightness,

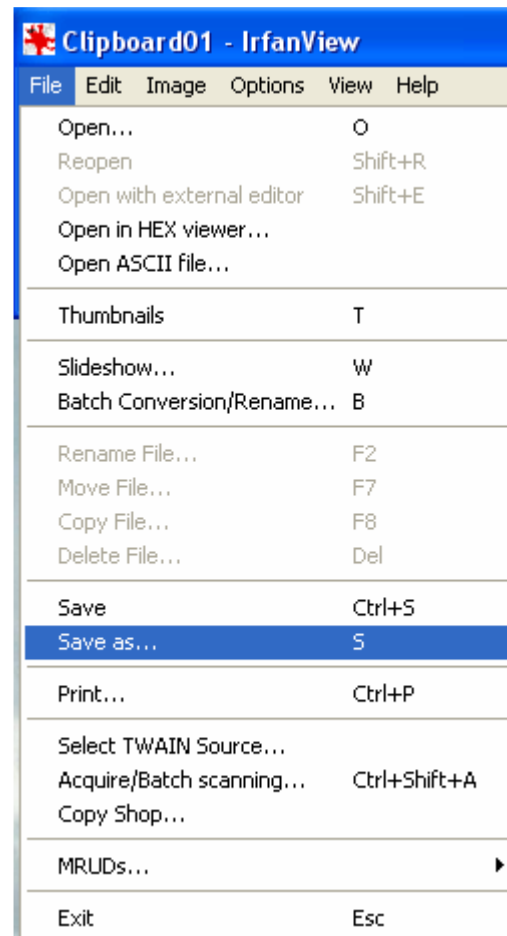contrast, colour balance, red eye reduction, image rotations, and many others.

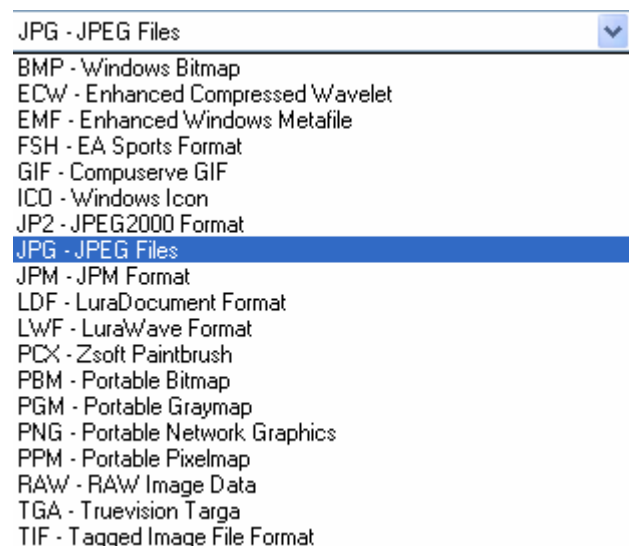Fig. 2. File-Save as option for formula (1)

Fig. 3. Graphical formats available in the current configuration

With this approach, any part of a document can be converted to a graphical file. This, of course, only needs to be done for the parts that do not have

their equivalent in www languages (HTML, XML, …).

In the following section, we assume that HTML statements for preparing web pages are well-known and only touch a specific problem for layouts with many frames.

3 Frames and Multiple Redraw

Within the framework of one project solved by Masaryk University, we faced the following problem. An Internet application is divided into four frames as shown in Fig. 4. Frame 1 only shows logos. Frame 2 includes names of several courses, with one of them selected and its contents displayed in Frame 3. The content of the item selected in this frame is then displayed in Frame 4. This last text may be scrolled in vertical direction and contains many bookmarks for displaying related information (again in Frame 4). The items in Frame 3 may be dynamically expanded and collapsed using the + and – signs by the Windows standard.
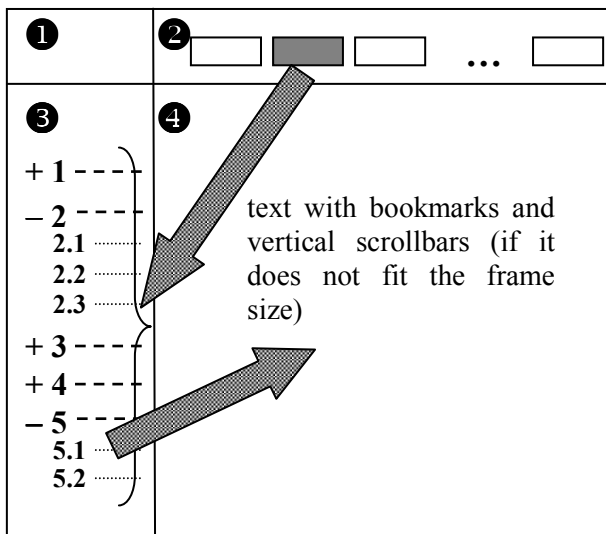

Fig. 4. Layout of a www application

The most difficult problem was how to redraw Frames 3 and 4 if a different course was selected in Frame 2. (We assumed that, after changing a course, the text in Frame 4 corresponded to the first item of the table of contents in Frame 3.) This means that, using a single click, we needed to redraw the contents of two frames at the same time.

After the necessary conversion from LATEX and Microsoft Word into the .HTML format, all the prepared texts were built into an interactive Internet application. Besides web sites, this application can also be run from local disks, CD-ROM or portable disks. It is sufficient to store all files while

respecting the directory structure according to the outline in Fig. 5. All the addresses in files are relative and thus the location of the root directory is irrelevant.

Fig. 5. Directory structure

The index.html root file has the following form:

```
<html>
    <title>FRVŠ-ESF</title>
    <frameset rows="54,*">
      <frameset cols="200,*">
        <frame src=".\Frames\F1_logo.html"
name="Logo" marginwidth="10" marginheight="10"
scrolling="auto" frameborder="0">
        <frame src=".\Frames\F2_menu.html"
name="Menu" marginwidth="10" marginheight="10"
scrolling="auto" frameborder="0">
      </frameset>
        <frame src=".\Frames\F3_Course_I.html"
name="Contents"  marginwidth="10"
marginheight="10" scrolling="auto"
frameborder="0">
    </frameset>
</html>
```

The key role is played by Frame 2 and the F2_menu.html file in the Frames directory. Items in F2_menu.html are inserted into a table, whose cells contain references to other frame files by the following pattern:

```
<td title="Jindřiška Šedová">
    <a href="F3_Course_I target="Contents">
    <b>Commercial Law </b></a>
</td>
```

Then, in the F3_Course_I.html frame file, the first chapter entitled ..\Course I\CI_P1.html from the course selected is mapped to Frame 4 by the following statements:

```
…
<tr><td nowrap style="cursor:hand;"
        onclick="display_m(1)">
    <img align="bottom" id="i1" src="plus.gif"
        width="16" height="16" alt="" border="0">
    </td>
```

```
    <td width="100%" nowrap valign="middle"
        style="cursor:hand;"
        onclick=" display_m (1)">1 <a
        href="..\Course I\CI_P1.html"
        target="F4_Texts"><b>Legal aspects of
business in the Czech Republic</b></a>
    </td>
</tr>
…
```

Here, display_m() is implemented in JavaScript to control the dynamic changes in Frame 3.

4 PDF (and PS) Files

Another approach, frequently used for making the L_AT_EX and MS Word web documents, consists in converting them into the .pdf (or .ps) format and including references to them in HTML documents.

While MiKT$_E$X [6] contains its own tool for a conversion from T$_E$X documents to .pdf, to create a .pdf file from MS Word, mostly Adobe Acrobat [1] is used. However, in its professional version, this software is rather expensive (currently USD 449 and USD 159 for update). On the other hand, Acrobat Reader [2] developed for reading (not creating) .pdf files is free. While new versions of Acrobat Reader are frequent and immediately available, in practice, Adobe Acrobat versions are not so frequently updated. From this point of view, it seems to be a good strategy to use another software to create .pdf files.

One of the possible tools is pdfFactory [5]. It is a small and inexpensive program developed by FinePrint Software. Its size is 1.4 MB with price about 40 Euro. It works as a virtual printer but, instead of printing on paper, a .pdf file is created. In this file, we can insert bookmarks (see Fig. 6), page numbering, headers, footers, multiple configurations, rearrange documents, automated email and also set security options.

As compared with Adobe Acrobat, .pdf documents generated by pdfFactory have a substantially smaller size than those generated by Adobe Acrobat and this fact make it possible to accelerate downloads of documents.

For completeness, we must add that mathematicians frequently use .ps (PostScript) documents, which can be generated from MS DOS versions of T$_E$X programs. These files can be read, e.g., by the GSview program [4]. PostScript documents can also be generated from Windows

versions of T$_E$X programs such as MiKT$_E$X, but, whenever possible, we prefer a more modern .pdf format.
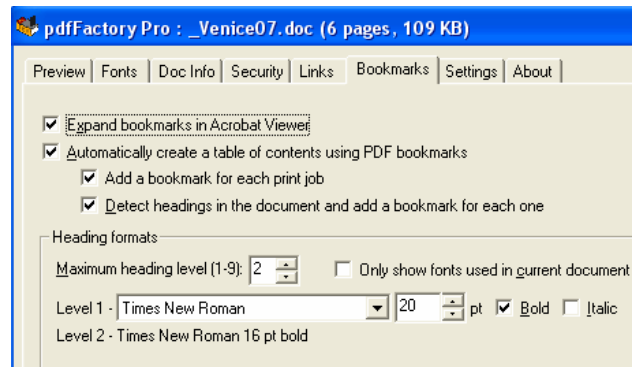

Fig. 6. Bookmarks in the pdfFactory

5 Conclusions

In the paper, several problems concerning the conversion of L_AT_EX and Word documents into the web environment were discussed. We showed that they can be resolved using only a very limited number of software tools, frequently inexpensive or even free, without the need of using specialized programmes such as CorellDraw or Adobe Illustrator for pictures, Adobe Acrobat for .pdf, etc.

References:
[1] Adobe Acrobat, http://www.adobe.com/products/acrobat/.
[2] Acrobat Reader, http://www.adobe.com/products/acrobat/readstep2.html.
[3] N. Drakos, The LaTeX2HTML Translator, http://www.rzg.mpg.de/from_external/GNU_docs/latex2html/index.html.
[4] GSview, http://pages.cs.wisc.edu/~ghost/gsview/.
[5] pdfFactory (create PDF files without Acrobat), http://www.pdffactory.com/.
[6] MiKT$_E$X … typesetting beautiful documents, http://www.miktex.org/.
[7] I. Skiljan, IrfanView, Freeware Graphic Viewer for Windows, http://www.irfanview.com/.

[8] L. Ullman, *PHP and MySQL for Dynamic Web Sites: Visual QuickPro Guide*, Peachpit Press, Berkeley, 2005.

[9] L. Welling and L. Thomson, *PHP and MySQL Web Development*, Sams Technical Publishing, Indianapolis, 2007.