

## **МОДУЛ ЗА LATEX ЕКСПОРТ В РАЗПРЕДЕЛЕНАТА ПЛАТФОРМА ЗА ЕЛЕКТРОННО ОБУЧЕНИЕ DISPEL**

**Ангел Голев, Николай Павлов,  
Георги Спасов, Кремена Стефанова**

***Резюме.** В рамките на разпределената платформа за електронно обучение (DisPeL) е създаден модул за експорт на електронните учебници, използвани в системата, към LaTeX формат. Целта е бързата подготовка на учебниците за печат на хартиен носител. Платформата разполага със свой редактор за въвеждане на електронните учебници, които се съхраняват в HTML формат със собствени стилове и структура. Създадени са HTML и LaTeX шаблони с еднозначно съответствие по между си, които описват структурата на електронния учебник и позволяват трансфера на съдържанието на учебника към LaTeX документ.*

**Ключови думи:** електронно обучение *DisPeL*, конвертиране на форматиран текст, *HTML*, *LaTeX*

### **1. ВЪВЕДЕНИЕ**

Учебното съдържание в разпределената платформа *DisPeL* [1, 2] се предоставя под формата на електронни учебници. Всеки един учебник се състои от предварително определен брой глави, които зависят от конкретната дисциплина и времето за изучаването ѝ. Достъпът до главите на електронния учебник става последователно, като за преминаването към следващата глава обучаемият трябва да реши съответния тест.

Съдържанието на електронния учебник се съхранява в системата в *HTML* формат, с използването на собствени стилове за отделните елементи на учебника. Целта е обучаващите се да могат да изучават електронния учебник, с помощта на браузър на устройства с различни размери на екрана.

Освен, че потребителите могат да си служат с електронното съдържание онлайн, преди започването на обучението учебниците се издават и на хартиен носител. В повечето случаи учебникът първо е подготвен на *Microsoft Word* и след това се въвежда в системата със специализиран редактор. Обикновено по-голямата част от текстовото съдържание се копира и само за някои части е

необходимо допълнително форматиране. Въвеждането не е бавно и сложно и може да се направи и от помощни лица.

Понякога съдържанието на електронния учебник се създава първо с редактора на *DisPeL*. Редакторът на системата не поддържа експорт към различни текстови формати и това създава трудности, ако учебникът трябва да се издаде на хартиен носител. Тази операция е доста трудоемка, независимо от това, че основния текст няма да се набира наново и от това каква друга текстообработваща програма ще използваме.

В рамките на системата за обучение е реализиран модул, който експортира съдържанието на електронния учебник в *LaTeX* документ. Впоследствие операциите по форматирането и странирането стават бързо и това ще ускори подготовката за издаване.

## 2. ШАБЛОНИ ЗА СЪОТВЕТСТВИЕ НА HTML КЪМ LATEX

Благодарение на това, че *DisPeL* съхранява електронните учебници в добре структурирани *HTML* документи, те могат напълно еднозначно да бъдат експортирани до *LaTeX* документи. Всички таговете в *HTML* намират съответствие в специфичните *LaTeX* среди и команди. Таговете имат точно определено начало, край, атрибути, съдържание и това позволява лесното свързване с елементите на *LaTeX*.

Стандартният *LaTeX* документ следва общата структура:

```
\documentclass[12pt]{article}
\usepackage[cp1251]{inputenc}
\usepackage[english,bulgarian]{babel}
\begin{document}
...
\end{document}
```

*Пример 1. Структура на LaTeX документ*

Всичко, което е поместено между `\begin{document}` и `\end{document}`, представлява тялото на *LaTeX* документа.

За всеки *HTML* документ трябва да се създаде общо описание на използваните конструкции, техните атрибути и съдържание. Всяка конструкция се описва с помощта на шаблон, който включва специални променливи. Специалните променливи намират приложение за означаване на местата, където ще бъде попълнено съдържанието от *HTML* документа.

Ще представим конкретни примери за някои от най-често използваните конструкции в *HTML* документите и тяхната реализация в *LaTeX*.

Специалните променливи в *HTML* документите, които се експортират до *LaTeX*, са обособени в скоби:

```
(text_content)
```

*Пример 2. Специална променлива в HTML документ*

Символът # се използва като префикс за променливите в *LaTeX* документа. Тези променливи имат вида:

#text_content#
----------------

**Пример 3. Специална променлива в *LaTeX* документ**

Пример 4 представя шаблон за експорт на заглавия, които не са номерирани. В случай, че се изисква номерация, тогава в съответната *LaTeX* команда се пропуска символа \*.

<i>HTML реализация</i>	<i>LaTeX реализация</i>
<h1> (H1_heading) </h1>	\section*{#H1_heading#}
<h2> (H2_heading) </h2>	\subsection*{#H2_heading#}
<h3> (H3_heading) </h3>	\subsubsection*{#H3_heading#}

**Пример 4. Експорт на заглавия**

В Пример 5 са илюстрирани *HTML* конструкциите за създаване на параграфи в съответствие с различните атрибути на тага <p>, които влияят на подравняването на текста. Също така е посочен *LaTeX* кода за всяка от тях.

<i>HTML реализация</i>	<i>LaTeX реализация</i>
<p align='justify'> (justified_paragraph_text) </p>	#justified_paragraph_text#
<p align='right'> (right_paragraph_text) </p>	\begin{flushright} #right_paragraph_text# \end{flushright}
<p align='center'> (centered_paragraph_text) </p>	\begin{center} #centered_paragraph_text# \end{center}
<p align='left'> (left_paragraph_text) </p>	\begin{flushleft} #left_paragraph_text# \end{flushleft}

**Пример 5. Експорт на параграфи**

В Пример 6 са предложени шаблони за стилове на шрифта, както и някои често използвани команди, като прекъсване на ред и създаване на коментар.

<i>HTML реализация</i>	<i>LaTeX реализация</i>
<b> (bold_text) </b>	\textbf{#bold_text#}
<em> (emp_text) </em>	\emph{#emp_text#}
<i> (italic_text) </i>	\textit{#italic_text#}
<small> (small_text) </small>	{\small #small_text#}
<strong> (strong_text) </strong>	\textbf{#strong_text#}
 	\newline
<!-- Comment goes here -->	%Comment goes here

**Пример 6. Експорт на стилове на шрифт и често използвани команди**

Пример 7 представя реализация на подредени и не подредени списъци.

<i>HTML</i> реализация	<i>LaTeX</i> реализация
<pre>&lt;ol&gt;   &lt;li&gt; (numbered_list_element) &lt;/li&gt; &lt;/ol&gt;</pre>	<pre>\begin{enumerate}   \item #numbered_list_element# \end{enumerate}</pre>
<pre>&lt;ul&gt;   &lt;li&gt; (bullet_list_element) &lt;/li&gt; &lt;/ul&gt;</pre>	<pre>\begin{itemize}   \item #bullet_list_element# \end{itemize}</pre>

Пример 7. Експорт на списъци

Ако е необходима модификация на водещия символ, *LaTeX* предоставя възможност за това чрез нейното задаване непосредствено след командата за създаване на елемент в списъка. По аналогичен начин се осъществява модификация при подреден списък.

<i>HTML</i> реализация	<i>LaTeX</i> реализация
<pre>&lt;li type="square"&gt;   (bullet_list_element) &lt;/li&gt;</pre>	<pre>\item[<math>\square</math>] #bullet_list_element#</pre>

Пример 8. Модификация на водещ символ

Пример 9 посочва конструкциите за създаване на таблица.

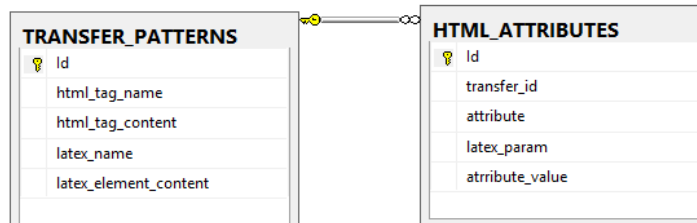
<i>HTML</i> реализация	<i>LaTeX</i> реализация
<pre>&lt;table&gt;   &lt;tr&gt;     &lt;td&gt; (row_1 element_1) &lt;/td&gt;     &lt;td&gt; (row_1 element_2) &lt;/td&gt;   &lt;/tr&gt;   &lt;tr&gt;     &lt;td&gt; (row_2 element_1) &lt;/td&gt;     &lt;td&gt; (row_2 element_2) &lt;/td&gt;   &lt;/tr&gt; &lt;/table&gt;</pre>	<pre>\begin{table}[h] \begin{tabular}{ c c } \hline #row_1 element_1# &amp; #row_1 element_2# &amp; \\ \hline #row_2 element_1# &amp; #row_2 element_2# &amp; \\ \hline \end{tabular} \end{table}</pre>

Пример 9. Експорт на таблица

След експортиране на *HTML* документа до *LaTeX* документ крайният файл може да бъде компилиран до *Portable Document Format (PDF)* формат.

### 3. МОДУЛ ЗА ЕКСПОРТ КЪМ LATEX

В базата от данни на *DisPeL* за съхранение на шаблоните за трансфер използваме таблиците *TRANSFER\_PATTERNS* и *HTML\_ATTRIBUTES*:



Част от шаблоните, описващи основните елементи на *HTML*, са предварително въведени в базата. Потребителите на системата могат да добавят нови шаблони за трансфер от *HTML* към *LaTeX* с опростен специализиран редактор.

The screenshot shows a web browser window with the address bar displaying 'localhost:2390/AddPattern.aspx'. The page title is 'Add new pattern'. There are four text input fields stacked vertically. The first is labeled 'HTML tag name:' and contains 'ol'. The second is labeled 'HTML element content:' and contains the HTML code '<ol> {ol\_content} </ol>'. The third is labeled 'LATEX tag name:' and contains 'enumerate'. The fourth is labeled 'LATEX element content:' and contains the LaTeX code '\\begin{enumerate} #ol\_content# \\end{enumerate}'. At the bottom of the form is a button labeled 'Add pattern'.

Модулът за експорт е написан на *C#*. Програмата използва готов *HTML* парсер [4], който отделя таговете на *HTML* съдържанието.

За всеки намерен таг се търси шаблон в таблицата *TRANSFER\_PATTERNS*. Ако бъде намерен, променливите на *LaTeX* шаблона се заместват с конкретно съдържание и резултата се добавя към файла с *LaTeX* документа.

Ако за даден таг не се намери запис в таблицата *TRANSFER\_PATTERNS*, към *LaTeX* документа се добавя коментар, включващ и целия оригинален *HTML* таг. Това може да се случи за неописани или много специфични елементи. Впоследствие това съдържание ще бъде форматирано ръчно.

Основната функция на модула е рекурсивна, като съдържанието на всеки таг се обработва с парсера и наличието на вложени тагове се обработва аналогично. Преведеното към *LaTeX* съдържание на текущия таг се включва в съдържанието на обхващания го.

```
static string GetInnerText(HtmlNode node, StringBuilder sb = null)
{
    if (sb == null)
    {
        sb = new StringBuilder();
    }
    foreach (var n in node.ChildNodes)
    {
        if (n.Name == "#text")
        {

```

```

    // GetTextFromNode - конвертира специалните символи
    sb.Append(GetTextFromNode(n));
    continue;
}
bool hasAttributes = false;
string attributes = String.Empty;
LatexParameters lp = new LatexParameters();
LatexPatterns lptr = new LatexPatterns();
if (LatexParameters.HasAttributes(n.Name) && n.HasAttributes)
{
    foreach (var attr in n.Attributes)
    {
        if (attributes.Contains("#content#"))
        {
            attributes = attributes.Replace("#content#", lp[attr]);
        }
        else
        {
            attributes = lp[attr];
        }
    }
    hasAttributes = !String.IsNullOrEmpty(attributes);
}

if (hasAttributes)
{
    string[] attributesArr = attributes.Split(new string[]
        { "#content#" }, StringSplitOptions.RemoveEmptyEntries);
    if (attributesArr.Length == 2)
    {
        sb.Append(attributesArr[0]);
    }
    // функцията FillPattern създава LaTeX елемента и ако в текста
    // има вложени тагове, тя се обръща отново към GetInnerText()
    FillPattern(lptr, sb, n); //
    if (attributesArr.Length == 2)
    {
        sb.Append(attributesArr[1]);
    }
}
else
{
    FillPattern(lptr, sb, n); //
}
}
return sb.ToString();
}

```

#### 4. ЗАКЛЮЧЕНИЕ

В рамките на разпределената платформа за електронно обучение *DisPeL* е създаден модул за трансфер на електронните учебници, използвани в системата, от *HTML* формат в *LaTeX* документ. Целта е бързото отпечатване на учебниците на хартиен носител. Шаблоните за *HTML/LaTeX* експорт, описващи основните елементи на електронните учебници, са въведени предварително, като в последствие могат да бъдат допълвани в зависимост от сложността на учебниците.

*Възможности за бъдещо развитие.* С реализираното взаимно еднозначно съответствие между *HTML* и *LaTeX* елементите, в системата лесно може да се осъществи и импорт от *LaTeX* документи, но това може да е ефективно само, ако има повече потребители, пишещи директно на *LaTeX*.

Модулът може да се разшири и с трансфер към други файлови формати. Например, могат да се направят шаблони за еднопосочно съответствие на *HTML* елементите към *Microsoft Word OpenXML* формат. Документите лесно могат да се структурират с помощта на *OpenXML SDK*. Трансферът лесно може да се разшири и към някои от форматите за електронни книги, например към *epub* или *mobi*.

#### БЛАГОДАРНОСТИ

Настоящата статия е частично финансирана по проект НИ13 ФМИ-002 към звено „Научна и приложна дейност” на ПУ „П. Хилендарски”.

#### ЛИТЕРАТУРА

- [1] Rahnev, A., Hr. Krushkov, N. Pavlov, M. Steiger and N. Valchanov, New master program “Business Software Technologies” in faculty of mathematics and informatics of Plovdiv university, *Proceedings of “Traditions, Direction, Challenges” – Jubilee National Scientific Conference with International Participation*, Smolyan, 2012, 173–176.
- [2] Rahnev, A., N. Pavlov and V. Kyurkchiev, Distributed Platform for e-Learning – DisPeL, *European International Journal of Science and Technology (EIJST)*, Vol. 3, No. 1, 2014, 95–109, ISSN: 2304-9693.
- [3] Rahnev, A., N. Pavlov, A. Golev, M. Stieger and T. Gardjeva, New Electronic Education Services Using the Distributed E-Learning Platform (DisPeL), *International Electronic Journal of Pure and Applied Mathematics (IEJPAM)*, Vol. 7, No. 2, 2014, 63–71, ISSN: 1314-0744.
- [4] Html Agility Pack, <http://htmlagilitypack.codeplex.com/>
- [5] <http://www.htmlgoodies.com/>
- [6] Dongen M., *LaTeX and Friends*, Springer, 2012.
- [7] Goossens M., S. Rahtz, *The LaTeX Web companion: integrating TeX, HTML, and XML*, Addison-Wesley, 2006.

Факултет по математика и информатика  
Пловдивски университет „Паисий Хилендарски“  
Бул. „България“ № 236, 4003 Пловдив, България  
[angelg@uni-plovdiv.bg](mailto:angelg@uni-plovdiv.bg), [nikolayp@uni-plovdiv.bg](mailto:nikolayp@uni-plovdiv.bg)  
[g.spasov@kodar.net](mailto:g.spasov@kodar.net), [kstefanova@uni-plovdiv.bg](mailto:kstefanova@uni-plovdiv.bg)

## **MODULE FOR LATEX EXPORT IN THE DISTRUBUTED PLATFORM FOR E-LEARNING DISPEL**

**Angel Golev, Nikolay Pavlov, Georgi Spasov, Kremena Stefanova**

***Abstract.** Within the distributed e-learning platform (DisPeL) it is created a module for exporting of electronic textbooks used in the system, to LaTeX format. The aim is the preparation of textbooks for printing on paper. The platform has its own editor to input electronic textbooks, which are stored in HTML format with own styles and structures. Corresponding HTML and LaTeX templates are created that describe the structure of the electronic textbook and allow the transfer of the contents of the textbook to LaTeX document.*