

Teoria Kompilacji i Kompilatory

Konwersja z języka DokuWiki -> HTML

Projekt zaliczeniowy

Wiktor Mendalka

Jakub Komnata

Adrian Kozień

Aleksander Janur




Spis treści

1. Zarys i cel projektu
2. Zestawienie obsługiwanych tagów Dokuwiki z ich odpowiednikami w HTML
3. Przykład tworzenia tabeli
4. Gramatyka
5. Scanner
6. Lista tokenów

Zarys i cel projektu

DokuWiki to środowisko umożliwiające tworzenie stron i współpracę przy ich budowie. Dane przechowywane są w plikach tekstowych, dzięki czemu nie potrzebna jest baza danych. DokuWiki zostało napisane w 2004 przez Andreas'a Gohr'a w PHP, co daje możliwości edycji stron, lecz projekt wykonany został w języku Java co pozwala na uruchomienie go na każdej platformie. W dobie rozwoju internetu i przemysłu każda firma musi posiadać swoją stronę internetową by zaistnieć w sieci, co przekłada się na jej reklamę i w konsekwencji zyski. Często firmy te nie mają styczności z projektowaniem i tworzeniem stron internetowych. Głównym celem projektu jest konwersja z DokuWiki na HTML. Narzędzie przez nas opracowane konwertujące tagi może okazać się pomocne dla ludzi bez większych zdolności informatycznych. Motywacją do realizacji tego projektu był fakt, że coraz częściej dostępne są programy pozwalające na tworzenie stron internetowych dla ludzi o małych zdolnościach informatycznych.

Zestawienie wszystkich obsługiwanych tagów w Docuwiki oraz odpowiedników w HTML

DocuWiki Tag	HTML Tag	Przykład	Opis słowny
bold	bold	bold	pogrubienie
<i>//italic//</i>	<i><i>italic</i></i>	<i>italic</i>	kursywa
<u>__underline__</u>	<u><u>underline</u></u>	<u>underline</u>	podkreślenie
_{subscript}	_{subscript}	subscript	indeks dolny
^{superscript}	^{superscript}	superscript	indeks górny
deleted	deleted	deleted	przekreślenie
\\	 		nowa linia
[[address description]]	des	description	hiperłącze
[[address {{wiki:imgname}}]]			obraz z hiperłączem
((footnote))	[1]	¹ footnote	przypis
==== Headline Level 3 ====	<h3>Headline Level 3</h3>	Headline lvl 3	nagłówek 3
=== Headline Level 4 ===	<h4>Headline Level 4</h4>	Headline lvl 4	nagłówek 4
== Headline Level 5 ==	<h5>Headline Level 5</h5>	Headline lvl 5	nagłówek 5
{{wiki:dokuwiki-128.png?200x50}}			obraz o zmienionej rozdzielczości
{{video.mp4 A funny video}}	<video width="320" height="240" controls> <source src="movie.mp4" type="video/mp4"> </video>		dodanie wideo mp4
* This is a list * The second * Different levels * Another item	 This is a list The second Different levels 	This is a list The second Different levels Another item	lista wypunktowana

	Another item 		
- This is a list - Another item - Deeper lvel - That's it	 This is a list The second < ol > Different levels </ ol > Another item </ ol >	The same list but ordered Another item Just use indention for deeper levels That's it	lista numerowana
8-)	😎	😏	😏
8-O	😳	🙄	🙄
:-(😕	😞	😞
:-)	😀	😊	😊
:-/	😒	😏	😏
:-\	😑	😏	😏
:-?	😕	😏	😏
:-D	😃	😄	😄
:-P	😋	😄	😄
:-O	😄	😄	😄
:-X	😷	😄	😄
:-	😐	😏	😏
;-)	😉	😊	😊
^ ^ _		😄	😄
:?:		?	?
!:		!	!
LOL			😄
FIXME			🔧 Fix Me!
DELETEME			🗑 Delete!
->	→		→
<-	←		←
<->	↔		↔
=>	⇒		⇒
<=	⇐		⇐
<=>	⇔		⇔
>>	»		»
<<	«		«

--	–		—
---	―		—
(c)	―		©
(tm)	―		™
(r)	®		®
"He thought..."	\“He thought ...\”	"He thought..."	"He thought..."
sentence 0 > sentence 1 >> sentence 2 > sentence 1 >> sentence 2 >>> sentence 3	sentence 0 | sentence 1 | | sentence 2 | sentence 1 | | sentence 2 | | | sentence 3	sentence 0 sentence 1 sentence 2 sentence 1 sentence 2 sentence 3	cytowanie (dialog)
<nowiki> text without conversation like this: http://www.google.pl</nowiki>	text without conversation like this: http://www.google.pl	text without conversation like this: http://www.google.pl	Tekst który nie jest podawany żadnej konwersji
<code>Here you can write preformatted code with all spaces </code>	<code> Here you can write preformatted code with all spaces </code>	Here you can write preformatted code with all spaces	Wpisanie tekstu z uwzględnieniem wszystkich białych spacji.
<code java>some code </code>	<pre name="code" class="java "> some code here </pre>	Some code	Kolorowanie składni dla języka Java
<code c_cpp>some code </code>	<pre name="code" class="c_cpp "> some code here </pre>	Some code	Kolorowanie składni dla języka C/C++
<code c_sharp>some code </code>	<pre name="code" class="c_sharp "> some code here </pre>	Some code	Kolorowanie składni dla języka C#
<code c-object>some code </code>	<pre name="code" class="c-object"> some code here </pre>	Some code	Kolorowanie składni dla języka Objective-C
<code php>some code </code>	<pre name="code" class="php"> some code here </pre>	Some code	Kolorowanie składni dla języka PHP

	some code here </pre>		
<file php myexmaple.php> <?php echo "hello world"; ?>	<dl class="file"> <dt> example.php</ dt> <dd> <pre class="code file php"> <? php echo "hello world"; ?> </pre> </dd></dl>		Kolorowanie składni z możliwością pobrania pliku z kodem źródłowym

Tworzenie tabel:

Przykładowa tabela

Heading 1	Heading 2	Heading 3
Row 1 Col 1	Row 1 Col 1	Row 1 Col 1
Row 2 Col 1	Some colspan	
Row 3 Col 1	Row 3 Col 2	Row 3 Col 3

Kod DokuWiki

```

^ Heading 1   ^ Heading 2   ^ Heading 3   ^
| Row 1 Col 1 | Row 1 Col 2 | Row 1 Col 3  |
| Row 2 Col 1 | some colspan (note the double pipe) ||
| Row 3 Col 1 | Row 3 Col 2 | Row 3 Col 3  |

```

Kod HTML:

```

<table border="1" style="border-collapse: collapse; border: 1px solid #ccc;">
<thead>
<tr style="background-color:#eee;">
<th>Heading 1</th>
<th>Heading 2</th>
<th>Heading 3</th>
</tr>
</thead>
<tr>
<td>Row 1 Col 1 </td>
<td>Row 1 Col 2 </td>
<td>Row 1 Col 3 </td>
</tr>
<tr>
<td>Row 2 Col 1 </td>
<td colspan="2">Row 2 Col 2 with colspan </td>

```

```

</tr>
<tr>
<td>Row 3 Col 1 </td>
<td>Row 3 Col 2 </td>
<td>Row 3 Col 3 </td>
</tr>
</table>

```

GRAMATYKA

```

Document = {whitespace}, {digit|unformattedText|formattedText|
programmingLanguageFormattedText|link|img|video|symbol|emoticon|table};

```

```

digit = ("1"|"2"|"3"|"4"|"5"|"6"|"7"|"8"|"9"|"0");

```

```

ascii_letter = ("a"|"b"|"c"|"d"|"e"|"f"|"g"|"h"|"i"|"j"|"k"|"l"|"m"|"n"|"o"|"p"|"q"|"r" |
"s"|"t"|"u"|"v"|"w"|"x"|"y"|"z"|"A"|"B"|"C"|"D"|"E"|"F"|"G"|"H"|"I"|"J"|"K"|"L" |
"M"|"N"|"O"|"P"|"Q"|"R"|"S"|"T"|"U"|"V"|"W"|"X"|"Y"|"Z");

```

```

keySymbol = ("!"|"@"|"#"|"$"|"%"|"^"|"&"|"("|")"|"{"|"}"|"["|"]"|","|"."|"'"|"<"
|">"|"~"|"`"|"+"|"_"|"*"|"\"|"\"\"|\"\"");

```

```

new_line = {"\\n"};

```

```

tabular = {"\\t"};

```

```

whitespace = {" " | new_line | tabular};

```

```

unformattedText = {digit|ASCII letter|whitespace|list};

```

```

Headline_1 = "=====",(unformattedText|formattedText),"====="

```

```

Headline_2 = "====", (unformattedText|formattedText),"===="

```

```

Headline_3 = "====", (unformattedText|formattedText),"===="

```

```

Headline_4 = "====", (unformattedText|formattedText),"===="

```

```

Headline_5 = "====", (unformattedText|formattedText),"===="

```

```

headline = Headline_1|Headline_2|Headline_3|Headline_4|Headline_5

```

```

boldSymbol = "***";

```

```

boldText = boldSymbol,(unformattedText|formattedText),boldSymbol;

```

```

italicSymbol = "//";

```

```

italicText = italicSymbol,(unformattedText|formattedText), italicSymbol;

```

```

underlineSymbol = "_";

```

```

underlineText = underlineSymbol,(unformattedText|formattedText),underlineSymbol;

```

```

subscriptStartSymbol = "<sub>";

```

```

subscriptEndSymbol = "</sub>";

```

```

subscriptText = subscriptStartSymbol,(unformattedText|formattedText),subscriptEndSymbol;

```

```

superscriptStartSymbol = "<sup>";

```

```

superscriptEndSymbol = "</sup>";

```

```

superscriptText = superscriptStartSymbol,(unformattedText|formattedText),superscriptEndSymbol;

```

```

deletedStartSymbol = "<del>";
deletedEndSymbol = "</del>";
deletedText = deletedStartSymbol,(unformattedText|formattedText),deletedEndSymbol;
footNoteStartSymbol = "({";
footNoteEndSymbol = "})";
footNote = footNoteStartSymbol,(unformattedText|formattedText),footNoteEndSymbol;
nowikiStartSymbol = "<nowiki>";
nowikiEndSymbol = "</nowiki>";
nowiki = nowikiStartSymbol,(unformattedText|formattedText),nowikiEndSymbol;

formattedText =
boldText|italicText|underlineText|subscriptText|superscriptText|deletedText|headline|footNote|nowiki

bulledList = "*", unformattedText, [contiuneBulledList];
continueBulledList = (bulledList |continueBulledList), new_line, [bulledList];
numberedList = "-", unformattedText;
contiuneNumberedList = (numberedList |contiuneNumberedList), new_line,
[numberedList];
list = {bulletedList| numberedList};

emot1 = "8-)";
emot2 = "8-O";
emot3 = ":-(";
emot4 = ":-)";
emot5 = "=)";
emot6 = ":-/";
emot7 = ":-\"";
emot8 = ":-?";
emot9 = ":-D";
emot10 = ":-P";
emot11 = ":-O";
emot12 = ":-X";
emot13 = ":-|";
emot14 = " ;-)";
emot15 = "^_^";
emot16 = "LOL";
emotQuestion = ":-?:";
emotExclamation = ":-!:";
emotFixMe = "FIXME";

```



```

emotDelete = "DELETEME";

emoticons =
{emot1,emot2,emot3,emot4,emot5,emot6,emot7,emot8,emot9,emot10,emot11,emot12,emot13,emot14,em
ot15,emot16,emotQuestion,emotExclamation,emotFixMe,emotDelete}

symb1 = "->";
symb2 = "<-";
symb3 = "<->";
symb4 = ">";
symb5 = "<=";
symb6 = "<=>";
symb7 = ">>";
symb8 = "<<";
symb8 = "--";
symb9 = "---";
symb10 = "(c)";
symb11 = "(tm)";
symb12 = "(r)";

symbol = {symb1,symb2,symb3,symb4,symb5,symb6,symb7,symb8,symb9,symb10,symb11,symb12}

linkStartSymbol = "[[";
linkEndSymbol = "]]";
linkSeparateSymbol = "|";

linkAllowChar = ( "-" | " " | "." | "~" | "!" | "*" | "'" | "(" | ")" | ";" | ":" | "@" | "&" | "=" | "+" | "$" | "," | "/"
| "?" | "%" | "#" | "[" | "]" );

addressText = {digit|ascii_letter|linkAllowChar}

address = "h","t","t","p",":","/","/",addressText,{".",addressText},{addressText,"/"}

link = linkStartSymbol,address,linkSeparateSymbol,(unformattedText|formattedText),linkEndSymbol;

imgStartSymbol = "[[";
imgEndSymbol = "]]";
imgSepAdrrlImgName = "|";
imgSeparateStartSymbol = "{{";
imgSeparateEndSymbol = "}}";

imgName = {digit|ascii_letter|linkAllowChar}

imgWikiName = "w","i","k","i",":",imgName;

img = imgStartSymbol, address, imgSepAdrrlImgName , imgSeparateStartSymbol , imgWikiName,
imgSeparateEndSymbol, imgEndSymbol;

videoStartSymbol = "{{";
videoEndSymbol = "}}";

```

```

videoSeparateSymbol = "|";
videoFileName = {digit|ascii_letter|linkAllowChar}
videoFilePath = videoFileName, ".", "m", "p", "4";
video = videoStartSymbol, videoFilePath, videoSeparateSymbol, unformattedText, videoEndSymbol;
programmingLanguageType = (""|"java"|"c_cpp"|"c_sharp"|"c-object"|"php");
startPart1Code = "<code ";
startPart2Code = ">";
endCode = "</code>";
code = startPart1Code, programmingLanguageType, startPart2Code, unformattedText, endCode;
tableHeaderSymbol = "^";
tableSimpleSymbol = "|";
tableAlignLeft = [whitespace], ({formattedText|unformattedText});
tableAlignCenter = whitespace, whitespace, whitespace, ({formattedText|unformattedText}),
whitespace, whitespace, whitespace;
tableAlignRight = whitespace, whitespace, whitespace, ({formattedText|unformattedText});
tableAlign = tableAlignLeft|tableAlignCenter|tableAlignRight;
tableHeader = {tableHeaderSymbol, tableAlign};
tableMergedRow = tableSimpleSymbol, tableAlign, tableSimpleSymbol;
tableSimpleRow = {(tableSimpleSymbol , tableAlign)|tableMergedRow };
tableRow = tableHeader, tableSimpleRow;
table = {tableRow, new_line};

```

SCANNER

Zadaniem skanera jest parsowanie tekstu - znalezienie symboli języka (atomów) oraz pominięcie znaków nie wpływających na działanie programu - np. komentarzy. Danymi wyjściowymi jest pośrednia postać programu źródłowego z opisami.

W naszym projekcie użyliśmy strukturę Enum z języka Java. Dzięki temu, gdy pojawi się dany symbol z rozpatrywanej gramatyki to zostaje przesłany do następnej części programu w celu weryfikacji. Aby pokazać działanie skanera, utworzono klasę, która na podstawie przesłanych enum'ów (symbolizujący tokeny) koloruje składnie.

Ponadto, wszystkie spacje, które są ignorowane według specyfikacji DokuWiki, także są obsługiwane przez nasz program.

LISTA TOKENÓW

TEXT, START_BOLD, BOLD, START_ITALIC, ITALIC, START_UNDERLINE, UNDERLINE, START_MONOSPACED, MONOSPACED, DEL, SUB, SUP, START_NEW_LINE, START_NEW_LINE_TAG, NEW_LINE_TAG, START_HYPERLINK, INNER_HYPERLINK, END_HYPERLINK, START_IMAGE, INNER_IMAGE, END_IMAGE, FOOTNOTE, START_FOOTNOTE, END_FOOTNOTE, START_LIST, IN_LIST, END_LIST, LIST_1_1, LIST_1_2, LIST_1_3, LIST_2_1, LIST_2_2, LIST_2_3, START_HEADLINE, START_HEADLINE5, START_HEADLINE4, START_HEADLINE3, HEADLINE4, HEADLINE3, HEADLINE5, TABLE_HEADING, START_TAG, IN_TAG, END_TAG, NOWIKI, CODE, SHOW, DONE

Uwaga: "TEXT" to zwykły tekst, zawierający liczby, litery i spacje nie będące w gramatyce DokuWiki