

Projekt zaliczeniowy z przedmiotu Teoria Kompilacji i Kompilatory

Kacper Tonia

Sławomir Kalandyk

Temat projektu

Tłumaczenie kodu źródłowego DokuWiki na język HTML

Instalacja i uruchomienie

Wymagane jest środowisko .NET Core 3.1. Po zainstalowaniu możliwe jest skompilowanie i uruchomienie aplikacji w następujący sposób:

cd DokuWikiTranslator.Cli/

dotnet run DokuWikiTranslator.Cli.csproj plik-wejscowy

lub

dotnet run DokuWikiTranslator.Cli.csproj plik-wejscowy > plik-wyjsciowy

Program wypisuje kod wynikowy na standardowe wyjście. Na standardowe wyjście błędu wypisywane są komunikaty o błędzie.

Zrealizowane funkcjonalności

1. Podstawowe formatowanie tekstu

DokuWiki supports **bold**, *italic*, underlined and monospaced texts. Of course you can combine all these.

You can use subscript and superscript, too.

Some ***mixed*** styling. You can mark something as ~~deleted~~ as well.

Proste operacje zapisywane w postaci “[znacznik otwierający] kod [znacznik zamykający]” zostały zrealizowane. Należą do nich m.in.: pogrubianie tekstu, podkreślenie, indeks dolny i górny itd. Znaczniki można mieszać, ale niepoprawna kolejność doprowadzi do błędu:

```
10 | You, a few seconds ago - Uncommi
11 | Some **mixed/** styling.// You can mar
12 |
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

Translation error at line 11:
Failed to parse token '**':
    Failed to parse token '//':
    Closing marker not found for '//'.

```

2. Nagłówki sekcji

```
==== Basic formatting ====
==== Subsection ====
== Subsubsection ==
DokuWiki ==== supports ==== bold,
```

Basic formatting

Subsection

Subsubsection

DokuWiki==== supports==== **bold**, *italic*, un

Znaczniki nagłówków działają tylko wtedy, gdy są umieszczone na początku linii.

3. Hiperłącza

Możliwe jest dodawanie linków za pomocą podwójnych nawiasów kwadratowych. Ze względu na problemy z parsowaniem adresów internetowych, zwłaszcza wykrywaniem końca takiego adresu, adresy nieoznaczone nawiasami kwadratowymi są traktowane jak zwykły tekst.

```
==== Hyperlinks ====
```

```
External links: www.example.com, [[http://example.com]] , [[example.com]]
```

```
Links with caption: [[www.example.com|Link caption]]
```

```
<h3> Hyperlinks</h3>
```

```
External links: www.example.com, <a href='http://example.com'>http://example.com</a> ,  
<a href='example.com'>example.com</a><br />
```

4. Wyłączenie formatowania

DokuWiki oferuje 2 znaczniki wyłączające formatowanie fragmentów kodu:

```
==== Disable formatting ====
```

```
<nowiki>This is some text which contains formatting, but nothing  
is done with it. The same is true for __this__ text//. </nowiki>
```

```
This is formatted.
```

```
%%More unformatted text: [[www.example.com]]%%
```

Disable formatting

This is some text which contains **formatting**, but nothing is done with it. The same is true for __this__ text//.

This **is formatted.**

More unformatted text: [[www.example.com]]

5. Przejście do nowej linii

Użycie pustej linii,
kombinacji “\\[spacja]” bądź
“\\[koniec linii]” powoduje
przejście do nowej linii.

```
==== Text flow ====
Paragraph.

Another paragraph.

This is some text with some linebreaks\\ Note that the
two backslashes are only recognized at the end of a line\\
or followed by\\ a whitespace \\this happens without it.
```

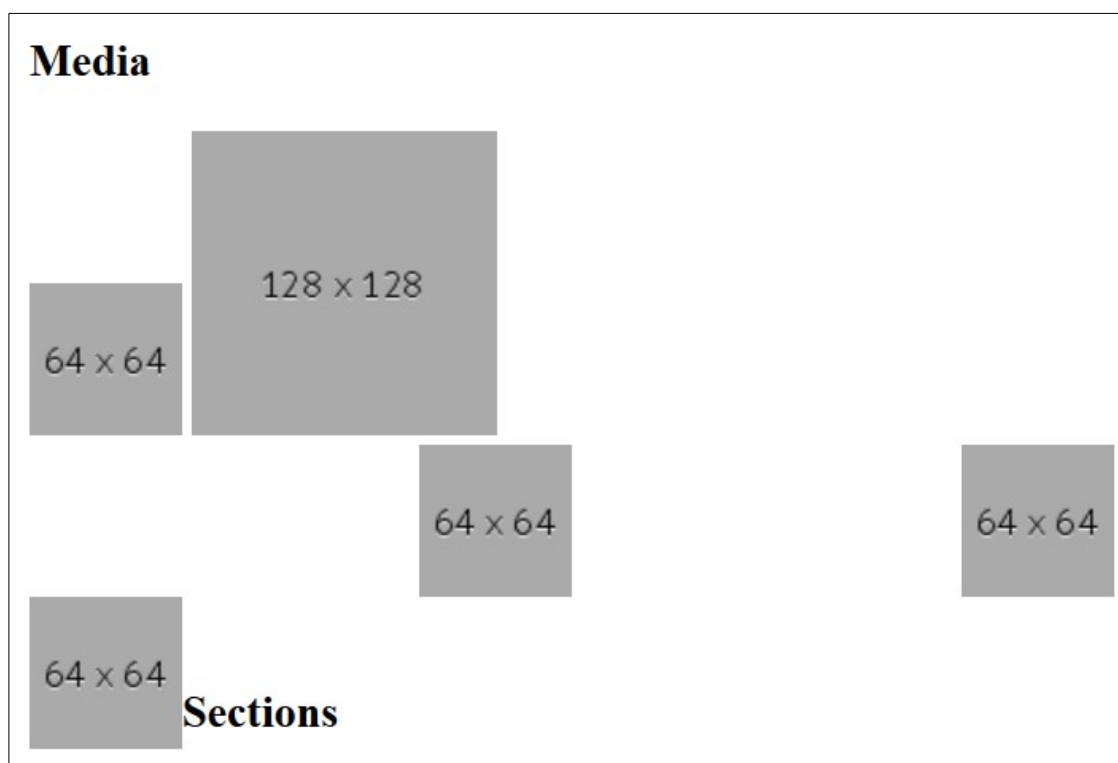
Paragraph.
Another paragraph.
This is some text with some linebreaks
Note that the two backslashes are only recognized at the end of a line
or followed by
a whitespace \\this happens without it.

6. Media

Dodanie obrazów do dokumentu odbywa się za pomocą zapisu:

- **{{adres}}**
- **{{adres|opis}}**

Możliwe jest wypozycjonowanie obrazka poprzez dodanie białych znaków przy lewym i/lub prawym znaczniku. Wykorzystane style są bardzo proste, więc układ tekstu może zostać zaburzony.



7. Separator

Ciąg 4 (lub więcej) myślników zostanie przekształcony na separator.

Some text.

Text after a separator.

8. Specjalne kombinacje znaków

Niektórym znakom specjalnym są przypisane odpowiednie kombinacje znaków.

Special: <-, <->, ->, =>, <=, <=>\\
(c) (tm)\\
<<cudzysłów niemiecki>>\\

Special: ←, ↔, →, ⇒, ⇐, ⇔
© TM
«cudzysłów niemiecki»

Architektura

Kod źródłowy podzielony jest na trzy projekty:

- Projekt **Application** zawiera faktyczną implementację skanera, parsera i generatora.
- Projekt **Cli** umożliwia uruchomienie i przeprowadzenie kompilacji z poziomu konsoli.
- Projekt **Application.Tests** zawiera testy jednostkowe.

Klasa **Translator** pozwala ukryć szczegóły tłumaczenia kodu za pojedynczym wywołaniem metody **Translate()**. Pod spodem, wykorzystywane są obiekty następujących klas:

- **Lexer** – odpowiada za wstępne skanowanie i rozbiecie kodu źródłowego DokuWiki na tokeny.
- **Parser** – buduje z ciągu tokenów drzewo obiektów implementujących interfejs **IDokuWikiTreeNode**.
- Generacja – każdy węzeł wspomnianego drzewa generuje odpowiadający mu obiekt implementujący interfejs **IHtmlSyntaxTreeNode**. Wreszcie, obiekty nowo powstałego drzewa generują kod wyjściowy w języku HTML.