## Grzegorz Bylina

## Tomasz Gajda

## Elpidiusz Wszołek

Translator Dokuwiki ⇒ HTML

Projekt z przedmiotu Języki Formalne i Kompilatory

# Wstęp

Celem projektu jest napisanie aplikacji przetwarzającej tekst sformatowany w składni Dokuwiki na tekst formatowany znacznikami HTML.

Oba języki są językami znacznikowymi, a więc wszystkie informacje oprócz czystego tekstu wyrażone są za pomocą wplecionych w niego znaczników. Składnia obu języków jest dość przejrzysta i niezbyt skomplikowana. Dokuwiki wydaje się być bardziej czytelna i zrozumiała szczególnie dla użytkowników nie przywykłych do czytania i analizy kodu w innych językach programowania, przy zachowaniu dużych możliwości formatowania tekstu oferowanych przez HTML.

Mimo pokrywających się zastosowań obu języków nie wszystkie możliwości Dokuwiki mają bezpośrednie odwzorowanie w HTML i odwrotnie. Przykładowo część możliwości Dokuwiki wspierana jest przez arkusz CSS, więc bez jego użycia nie da się uzyskać zadowalającego wyniku. HTML również posiada funkcjonalności, których wykorzystania nie przewidzieli twórcy Dokuwiki.

# Gramatyka (w rozszerzonej notacji BNF)

CYFRA: 0|1|2|3|4|5|6|7|8|9

LITERA : ( A|B|C|D | …) | ( a|b|c|d |…)

NUMER : {CYFRA}

ALNUM: CYFRA | LITERA

WORD: { ALNUM | OGONEK }

ENTER : "\r"| "\n"| "\r\n"

NOWALINIA : "\\"

INNY\_ZNAK: "~" | "`" | "!" | "@" | "#" | "$" | "%" | "&" | "(" | ")" | "\_" | "+" | "=" | "{" | "}" | "[" | "]" | ";" | ":" | "'" | "\"" | "\\" | "<" | ">" | "," | "." | "/" | "?" | "\*" | "-"

WCIĘCIE : SPACE SPACE

AKAPIT : ENTER ENTER

LISTAN : WCIĘCIE "\*"

LISTAP : WCIĘCIE "-"

BOLD : "\*\*"

ITALICS : "//"

UNDERLINE : "\_\_"

NO\_FORMAT : "%%"

MONO : "''"

SPACE : " " | "\t"

LINIA : "---" {"-"} {[SPACE]} ENTER

H1 : "======" {["="]}

H2 : "====="

H3 : "===="

H4 : "==="

H5 : "=="

HC : "==" {"="} ENTER

DOMENA : { ({ ALNUM} ".") } WORD

URL : ("www." | "http://") DOMENA ( "/" {[ WORD ]} {("." < WORD >)}

EMAIL : "<" EMAIL\_NAZWA "@" DOMENA

EMAIL\_NAZWA : { ({ ALNUM } ".") } { ALNUM }

WYMIARY : "?" NUMER "x" NUMER

EMOTKA: "8-)" | "8-O" | ":-(" | ":-)" | "=)" | ":-/" | ":-\\" | ":-?" | ":-D" | ":-P" | ":-O" | ":-X" | ":-|" | ";-)" | "^\_^" | ":?:" | ":!:" | "LOL" | "FIXME" | "DELETEME"

SYMBOL : "->" | "<-" | "<->" | "=>" | "<=" | "<=>" | "--" | "---" | "(c)" | "(tm)" | "(r)"

HEADER : "^"

KOLUMNA : "|"

VERT\_SPAN : ":::" zrobić albo wyrzucić stąd I nie wspominać

LINKO : "[["

LINKC : "]]"

OBRO : "{{"

OBRC : "}}"

DELO : "<del>"

DELC : "</del>"

SUPO : "<sup>"

SUPC : "</sup>"

SUBO : "<sub>"

SUBC : "</sub>"

CODEO : "<code>"

CODEC : "</code>"

QUOTE : ENTER {(">")}

FILEO : "<file>"

FILEC : "</file>"

NOWIKIO : "<nowiki>"

NOWIKIC : "</nowiki>"

FOOTNOTEO : "((" zrobić albo wyrzucić i zapomnieć

FOOTNOTEC : "))" zrobić albo wyrzucić i zapomnieć

SFORMATOWANY : FOOTNOTES | EMAIL | BOLDED | CZYSTYTEKST | ITALICS | UNDERLINED | MONOSPACED | NOFORMAT | NOWIKI | SUB | SUP | DEL | LINK | CODE | FILE

BOLDED : BOLD { SFORMATOWANY } BOLD

ITALICS : UNDERLINE { SFORMATOWANY } UNDERLINE

MONOSPACED : MONO { SFORMATOWANY } MONO

NOFORMAT : NO\_FORMAT { ANYTHING } NO\_FORMAT

NOWIKI : NO\_WIKIO { ANYTHING } NO\_WIKIC

CODE : CODEO { ANYTHING } CODEC

FILE : FILEO { ANYTHING } FILEC

SUB : SUBO { SFORMATOWANY } SUBC

SUP : SUPO { SFORMATOWANY } SUPC

DEL : DELO { SFORMATOWANY } DELC

HEAD1 : H1 { ANYTHING } HC

HEAD2 : H2 { ANYTHING } HC

HEAD3 : H3 { ANYTHING } HC

HEAD4 : H4 { ANYTHING } HC

HEAD5 : H5 { ANYTHING } HC

QUOTELINIA : QUOTE { SFORMATOWANY } [ ENTER ]

LISTAU : [WCIĘCIE] { LISTAELEMENT }

LISTAELEMENT : ( (LISTAN | LISTAP) { [SFORMATOWANY] } ENTER ) | (LISTA)

TABWIERSZ : ENTER { KOMORKA } { [(KOLUMNA | HEADER)] }

KOMORKA : (KOLUMNA | HEADER) { SFORMATOWANY } { [KOLUMNA] }

FOOTNOTES : FOOTNOTEO CZYSTYTEKST FOOTNOTEC

OBRAZEK : OBRO (URL | DOMENA | CZYSTYTEKST) OBRC

LINK : URL | (LINKO (URL | DOMENA | CZYSTYTEKST) ( [ KOLUMNA CZYSTYTEKST ] ) )

SLOWO : WORD | NUMER

CZYSTYTEKST : { ( SLOWO | INNY\_ZNAK | SPACE ) }