

**AKADEMIA GÓRNICZO-HUTNICZA**  
**im. S. Staszica w Krakowie**



*Wydział Elektrotechniki, Automatyki, Informatyki  
i Inżynierii Biomedycznej*

***“Translator LatexMath - MathML”***

---

**Kierunek:**

*Informatyka:*

**Przedmiot:**

*Teoria kompilacji i Kompilatory*

**Prowadzący:**

*dr. inż. Wojciech Szmuc*

**Autorzy:**

*Krzysztof Pala*

*Szymon Czaplak*

**Data utworzenia:**

*Semestr letni 2019/2020*

*Teoria kompilacji i kompilatory*  
**Translator LatexMath -> MathML**

<b>Wstęp</b>	<b>3</b>
Opis programu	3
Użyte technologie	3
Autorzy	3
<b>Opis projektu</b>	<b>4</b>
Skaner	4
Parser	5
Przykładowe drzewo składniowe	6
Translator	7
<b>Wynik programu</b>	<b>9</b>

# I. Wstęp

## Opis programu

Program umożliwia translację dokumentu w języku latex zawierającego różnego rodzaju wzory i symbole matematyczne zawarte w paczce amsmath na dokument html z wykorzystaniem języka MathML. Głównym jego aspektem jest tłumaczenie fragmentów matematycznych, więc elementy dokumentu latex, które nie są z tym ściśle związane są pomijane. Program sprawdza się w tłumaczeniu dokumentów latex zawierającym - oprócz wzorów i symboli matematycznych - podstawowe funkcje latex'a.

## Użyte technologie

Program został stworzony przy użyciu środowiska programistycznego **IntelliJ IDEA**. Do konstrukcji zagadnień związanych z *language recognition* została wykorzystana biblioteka **ANTLR** i **ANTLR-runtime**, a także plug-in'a ANTLR. Program stworzono w języku **Java**.

## Autorzy

Projekt został wykonany w roku akademickim 2019/2020 w ramach przedmiotu Teoria Kompilacji i Kompilatory na WEAIiB AGH im. S. Staszica przez **Szymona Czapłaka** oraz **Krzysztofa Pala**.

## II. Opis projektu

Przykładowe elementy komponentów programu.

### Skaner

W części 'Keywords' zawarte są elementy standardowego Latex'a.

```
// Keywords:
DOCUMENTCLASS: '\\documentclass';
PACKAGES: '\\usepackage';
BEGIN: '\\begin';
END: '\\end';
TITLE: '\\title';
AUTHOR: '\\author';
DATE: '\\date';
FRAC: '\\frac';
INT: '\\int';
MAKETITLE: '\\maketitle';
SECTION: '\\section';
LABEL: '\\label';
```

W dalszej części odbywa się grupowanie różnych wykorzystywanych później symboli matematycznych.

```
// Simple symbols
PLUS: '+';
MINUS: '-';
EQ: '=' | '&=';
FACT: '!';
SLASH: '/';
MUL: '*';
POWER: '^';
L: '<';
G: '>';
OPEN_PAREN: '(';
CLOSE_PAREN: ')';
OPEN_BRACE: '{';
CLOSE_BRACE: '}';
OPEN_BRACKET: '[';
CLOSE_BRACKET: ']';
ANDSGN: '&';
NEWLINE: '\\\\';
SQRT: '\\sqrt';
ULTNE: '\\cdot';
```

## Teoria kompilacji i kompilatory

### Translator LatexMath -> MathML

```
// Greek letters
ALPHA: '\\alpha';
BETA: '\\beta';
GAMMA: '\\gamma';
DELTA: '\\delta';
EPSILON: '\\epsilon';
OMEGA: '\\omega';
PI: '\\pi';

//Trigonometric functions
SIN: '\\sin';
COS: '\\cos';
TAN: '\\tan';
COT: '\\cot';
ARCSIN: '\\arcsin';
ARCCOS: '\\arccos';
ARCTAN: '\\arctan';
ARCCOT: '\\arccot';
```

Skaner ignoruje białe znaki oraz komentarze.

```
// Ignore:
COMMENT: '%' (STRING | LETTER | WS | '-' )+ EOL-> channel(HIDDEN);
EOL: '\n' -> channel(HIDDEN);
WS: ' ' -> channel(HIDDEN);
```

## Parser

W poniższym fragmencie parsera opisane są reguły dla standardowych latex'owych wyrażeń, a także reguły ściśle związane z latex math (obsługa formatów equation, matrix, inline itd.)

```
// Standard for tex documents
prule: DOCUMENTCLASS OPEN_BRACE CLASS CLOSE_BRACE packages data document;
data: title? author? date?;
packages: ((PACKAGES OPEN_BRACE PACKAGE CLOSE_BRACE) | PACKAGES OPEN_BRACKET (STRING | DECIMAL_INTEGER)+ CLOSE_BRACKET OPEN_BRACE PACKAGE CLOSE_BRACE)+;
document: BEGIN OPEN_BRACE DOCUMENT CLOSE_BRACE MAKETITLE? body END OPEN_BRACE DOCUMENT CLOSE_BRACE;
body: (equation | matrix | align | text | multiline | inline | centered | displaymath | section)+ ;
section: SECTION OPEN_BRACE (STRING | WS | LETTER)+ CLOSE_BRACE (equation | matrix | align | text | multiline | inline | centered | displaymath)+;

// Document data
title: TITLE OPEN_BRACE (STRING | WS)+ CLOSE_BRACE;
author: AUTHOR OPEN_BRACE STRING WS? STRING? CLOSE_BRACE;
date: DATE OPEN_BRACE (STRING | DECIMAL_INTEGER)+ CLOSE_BRACE;

// Different types of mathematical notations in amsmath
equation: BEGIN OPEN_BRACE EQUATION CLOSE_BRACE label? (math_formula NEWLINE)* math_formula END OPEN_BRACE EQUATION CLOSE_BRACE;
matrix: BEGIN OPEN_BRACE MATRIX CLOSE_BRACE matrix_formula END OPEN_BRACE MATRIX CLOSE_BRACE;
align: BEGIN OPEN_BRACE ALIGN CLOSE_BRACE (math_formula | NEWLINE)* END OPEN_BRACE ALIGN CLOSE_BRACE;
multiline: BEGIN OPEN_BRACE MULTILINE CLOSE_BRACE (math_formula | NEWLINE)* END OPEN_BRACE MULTILINE CLOSE_BRACE;
inline: (BACKSLASH OPEN_PAREN math_formula BACKSLASH CLOSE_PAREN) | (DOLLAR math_formula DOLLAR);
centered: BACKSLASH OPEN_BRACKET math_formula BACKSLASH CLOSE_BRACKET;
displaymath: BEGIN OPEN_BRACE DISPLAYMATH CLOSE_BRACE math_formula END OPEN_BRACE DISPLAYMATH CLOSE_BRACE;
```

Krzysztof Pala, Szymon Czaplak  
AGH, WEAIIB, Informatyka S6

## Teoria kompilacji i kompilatory

### Translator LatexMath -> MathML

Symbole matematyczne ze skanera pogrupowane są na kategorie. Ułatwia to ich późniejszą obsługę w translatorze.

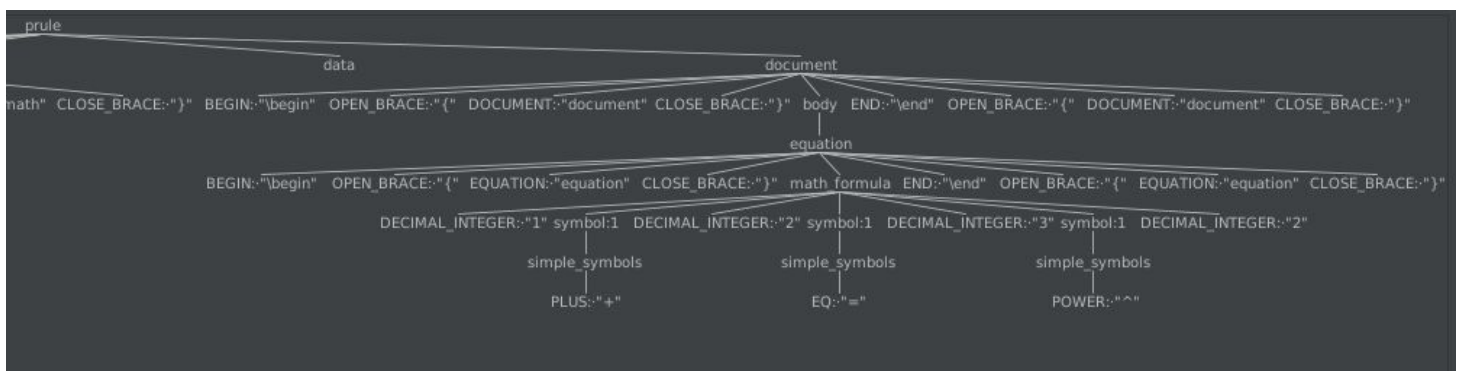
```
// Collections
simple_symbols: PLUS | MINUS | EQ | POWER | FACT | SLASH | MUL | L | G | OPEN_PAREN | CLOSE_PAREN | OPEN_BRACKET | CLOSE_BRACKET | ANDSGN | ULINE;
relation_symbols: LEQ | SUBSET | SUBSETEQ | GEQ | SUPSET | SUPSETEQ;
binary_operations: TIMES | DIV | AST | STAR;
logic_notation: EXISTS | NEXISTS | FORALL | NEG | IN | NIN | AND | OR | RIGHTARROW | LEFTARROW;
greek_letters: ALPHA | BETA | GAMMA | DELTA | EPSILON | OMEGA | PI;
trigonometric: SIN | COS | TAN | COT | ARCSIN | ARCCOS | ARCTAN | ARCCOT;
symbol: simple_symbols | relation_symbols | greek_letters | binary_operations | logic_notation | trigonometric;
keyword: CLASS | PACKAGE | DOCUMENT | EQUATION | ALIGN | MATRIX | MULTLINE | DISPLAYMATH;
```

Opisane są formuły dla różnych wyrażeń matematycznych.

```
// Formulas
math_formula: (DECIMAL_INTEGER | symbol | fraction | sqrt | integral | LETTER | STRING)+;
matrix_formula: DECIMAL_INTEGER | ((DECIMAL_INTEGER ANDSGN)+ DECIMAL_INTEGER NEWLINE)+ (DECIMAL_INTEGER ANDSGN)+ DECIMAL_INTEGER;
fraction: FRAC OPEN_BRACE math_formula CLOSE_BRACE OPEN_BRACE math_formula CLOSE_BRACE;
integral: INT ULINE LETTER POWER LETTER;
sqrt: Sqrt OPEN_BRACE math_formula CLOSE_BRACE;
```

### Przykładowe drzewo składniowe

Poniżej zaprezentowana jest część drzewa składniowego prostego dokumentu z jednym równaniem typu 'equation'.



```
\documentclass{article}
\usepackage{amsmath}
```

```
\begin{document}
```

```
\begin{equation}
1 + 2 = 3^2
\end{equation}
```

```
\end{document}
```

Krzysztof Pala, Szymon Czaplak  
**AGH, WEAIiB, Informatyka S6**

## Translator

Do wykonania translatora zostały wykorzystane klasy wygenerowane przez ANTLR. Została stworzona klasa Listener, która rozszerza wygenerowaną wcześniej klasę BaseListener.

```
public String getHtml() { return html; }

private void append(String s) { html = html + s; }

@Override
public void enterPrule(LatexParser.PruleContext ctx){
    html = html + "<!DOCTYPE html>\n<html>\n";
}

@Override
public void exitPrule(LatexParser.PruleContext ctx){
    html = html + "</html>";
}

@Override
public void enterData(LatexParser.DataContext ctx){ html = html + "<head>\n"; }

@Override
public void enterTitle(LatexParser.TitleContext ctx){
    html = html + "<title> ";
    StringBuilder title = new StringBuilder();
    int childCount = ctx.getChildCount();

    for(int i = 2; i<childCount-1; i++){
        title.append(ctx.getChild(i).getText()).append(" ");
    }

    html = html + title.toString() + " </title>\n";
}
```

```
@Override
public void enterMatrix(LatexParser.MatrixContext ctx){
    append("<math>\n<mrow>\n<mo>[</mo>\n<mtable>\n");
}

@Override
public void exitMatrix(LatexParser.MatrixContext ctx){
    append("</mtable>\n<mo>]</mo>\n</mrow>\n</math>\n");
}

@Override
public void enterMatrix_formula(LatexParser.Matrix_formulaContext ctx){
    append("<mtr>\n");
    for (int i=0; i<ctx.getChildCount(); i++){
        if(!ctx.getChild(i).getText().equals("&")){
            if(ctx.getChild(i).getText().equals("\\\\\\")){
                append("</mtr>\n<mtr>\n");
            }
            else {
                append("<mt><mn>" + ctx.getChild(i).getText() + "</mn></mt>\n");
            }
        }
    }
    append("</mtr>\n");
}
```



*Teoria kompilacji i kompilatory*  
**Translator LatexMath -> MathML**

```
@Override
public void enterMath_formula(LatexParser.Math_formulaContext ctx){
    boolean v1 = ctx.getParent().getClass().toString().equals("class gen.LatexParser$EquationContext");
    boolean v2 = ctx.getParent().getClass().toString().equals("class gen.LatexParser$FractionContext");

    if(v1 || v2) {
        append("<mathrow>\n");
    }
}

@Override
public void exitMath_formula(LatexParser.Math_formulaContext ctx){
    boolean v1 = ctx.getParent().getClass().toString().equals("class gen.LatexParser$EquationContext");
    boolean v2 = ctx.getParent().getClass().toString().equals("class gen.LatexParser$FractionContext");

    if(v1 || v2) {
        append("</mathrow>\n");
    }
}

@Override
public void enterSimple_symbols(LatexParser.Simple_symbolsContext ctx) {
    String symbol = ctx.getText();
    boolean isSet = false;

    if(symbol.equals("(")){html = html + "<mfenced open=\"(\" close=\">\")\">\n"; isSet=true;}
    if(symbol.equals(")")){html = html + "</mfenced>\n"; isSet=true;}

    if(symbol.equals("{")){html = html + "<mfenced open=\"{\" close=\">\"}\">\n"; isSet=true;}
    if(symbol.equals("}")){html = html + "</mfenced>\n"; isSet=true;}

    if(symbol.equals("[")){html = html + "<mfenced open=\"[\" close=\">\"]\">\n"; isSet=true;}
    if(symbol.equals("]")){html = html + "</mfenced>\n"; isSet=true;}

    if(symbol.equals("&")){isSet=true;}
}
```



## Wynik programu

Poniżej zaprezentowane są rezultaty translacji. Prezentowane są zarówno zmiany w kodzie jak i wyglądzie dokumentów.

### LATEX:

Amsmath example

ShareLaTeX Templates

June 2017

**1 Introduction**

$$(1) \quad A = \frac{\pi r^2}{2} = \frac{1}{2} \pi r^2$$

**2 Writing a single equations**

$$\pi i + 1 = 0 \quad (2)$$

The beautiful equation is known as the Euler equations

**3 Displaying long equations**

$$p(x) = 3x^6 + 14x^5y + 590x^4y^2 + 19x^3y^3 - 12x^2y^4 - 12x^5 + 2y^6 - a^3b^3$$

### HTML:

Amsmath example    ×    LatexMath to MathML    ×    +

← → ↺ ↻

file:///home/krzysztof/Dokumenty/Kompilatory/LatexToHtml/output.h

**Introduction**

$$A = \frac{\pi r^2}{2} = \frac{1}{2} \pi r^2$$

**Writing a single equations**

$$\pi i + 1 = 0$$

The beautiful equation is known as the Euler equations

**Displaying long equations**

$$px = 3x^6 + 14x^5y + 590x^4y^2 + 19x^3y^3 - 12x^2y^4 - 12x^5 + 2y^6 - a^3b^3$$

**Matrixes**

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

# Teoria kompilacji i kompilatory

## Translator LatexMath -> MathML

### LATEX:

```
\documentclass{article}
\usepackage{...}

\title{Amsmath example}
\author{ShareLaTeX Templates}
\date{June 2017}

\begin{document}

  \maketitle
  \section{Introduction}

  \begin{equation} \label{eq1}
    A \& = \frac{\pi}{r^2} \cdot 2 \quad \backslash \backslash
    \& = \frac{1}{2} \cdot \pi \cdot r^2
  \end{equation}

  \section{Writing a single equations}

  \begin{equation}
    \pi \cdot 1 + 1 = 0
  \end{equation}

  The beautiful equation is known as the Euler equations

  \section{Displaying long equations}

  \begin{multline*}
    p(x) = 3x^6 + 14x^5y + 590x^4y^2 + 19x^3y^3 \backslash \backslash
    - 12x^2y^4 - 12x^5 + 2y^6 - a^3b^3
  \end{multline*}

  \section{Matrixes}

  \begin{matrix}
    1 & \& 0 \backslash \backslash \\
    0 & \& 1
  \end{matrix}
```

### HTML:

```
<!DOCTYPE html>
<html>
<head>
<title> Amsmath example </title>
<meta name="author" content="ShareLaTeX Templates ">
<meta name="search_date" content="June 2017 ">
</head>

<body>
<section>
<h1> Introduction </h1>
<p>
<math>
<mstyle display="block">
<mrow>
<mn>A</mn>
<mo>=</mo>
<math xmlns='http://www.w3.org/1998/Math/MathML' display='block'>
<mfrac>
<mrow>
<mi>\pi</mi>
<msup>
<mn>r</mn>
</msup>
</mrow>
</mfrac>
</math>
</mrow>
<mrow>
<mo>=</mo>
<math xmlns='http://www.w3.org/1998/Math/MathML' display='block'>
<mfrac>
<mrow>
<mi>1</mi>
</mrow>
```

Krzysztof Pala, Szymon Czaplak  
AGH, WEAIIB, Informatyka S6