

# Parsing $\text{\LaTeX}$ for $\text{\LaTeX}$ file analysis and transformation

*Valentinas Kriaučiukas*

*Lukas Razinkovas*

*Lolita Žamoitinaitė*

**VTeX**

[www.vtex.lt](http://www.vtex.lt)

2015 April 29

2015 April 29

## My goals

When we need to parse  
 $\LaTeX$

Parsers versus  
converters

Our ongoing parser  
project

- Present our (VTeX) interests and experience in this area
- Specify the problem (what means to parse a  $\LaTeX$  code)
- Show use cases interesting for a wider audience
- Inspire developers' interest in solving the problem using  $\TeX$

My goals

When we need to parse  $\text{\LaTeX}$

Parsers versus converters

Our ongoing parser project

- In a conversion to XML
- In checking  $\text{\LaTeX}$  coding correctness, for example,
  - spaces after points
  - punctuation in math formulas
  - consistency per article
  - XML structure requirements
- In spell checking
- In data preparation for machine learning methods
- ...

My goals  
When we need to parse  
 $\LaTeX$

Parsers versus  
converters

Requirements for a  
converter

Requirements for a  
parser

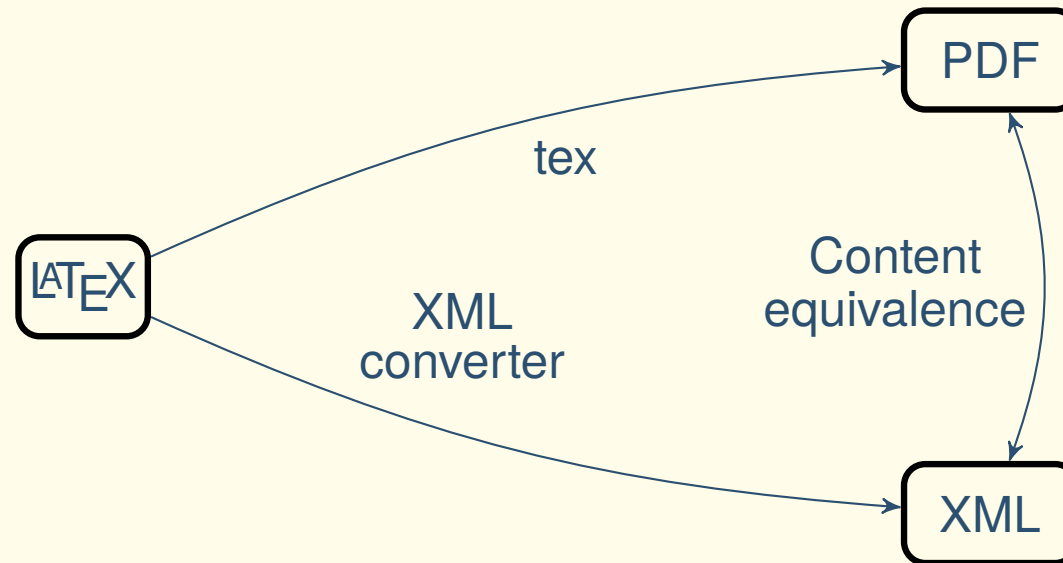
Possible branches of  
this talk

Our ongoing parser  
project

---

# Parsers versus converters

# Requirements for a converter



My goals  
When we need to parse  
L<sup>A</sup>T<sub>E</sub>X

Parsers versus  
converters

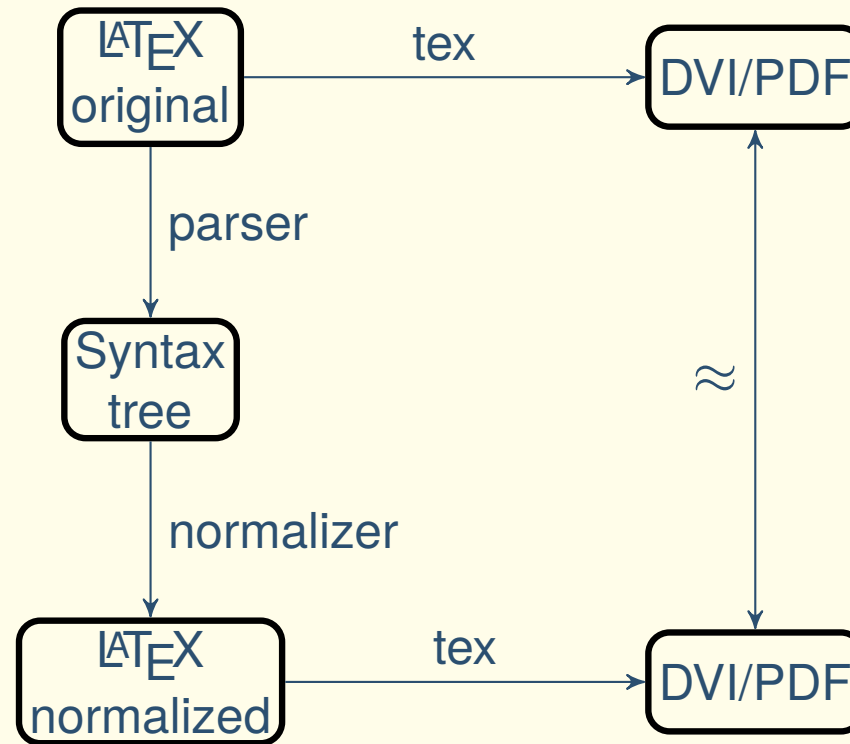
Requirements for a  
converter

Requirements for a  
parser

Possible branches of  
this talk

Our ongoing parser  
project

# Requirements for a parser



My goals  
When we need to parse  
LaTeX

Parsers versus  
converters

Requirements for a  
converter

Requirements for a  
parser

Possible branches of  
this talk

Our ongoing parser  
project

My goals

When we need to parse  
 $\LaTeX$

Parsers versus  
converters

Requirements for a  
converter  
Requirements for a  
parser

Possible branches of  
this talk

Our ongoing parser  
project

- $\LaTeX$  authors versus typesetters
  - Distinction of their tasks
  - Parser use in authoring tools
- The ongoing VTeX parser project
  - The current statistics
  - The conditions of the relative success
  - The  $\LaTeX$  syntax database
    - How it was initiated and is being filled in?
    - Why it is better than a built-in code?
    - How it could be opened, maintained and extended?
- Parsing versus file reading by  $\TeX$ 
  - How  $\TeX$  engine could help to parse  $\LaTeX$ ?
  - Would anybody of  $\TeX$  developers be interested to work in this direction?

My goals  
When we need to parse  
 $\text{\LaTeX}$

Parsers versus  
converters

---

**Our ongoing parser  
project**

The current parsing  
statistics

Scheme of the parser

To be continued...

# Our ongoing parser project



## The current parsing statistics

Characteristics	Number	(%)
Preparation failed	5	0.02
Prepared, compilation errors	2429	9.21
The rest is counted as 100%		
Success	19428	81.12
Undefined command	2029	8.47
Not implemented command	1031	4.30
Parsing failed	500	2.09
Normalization failed	12	0.05
Normalized compilation failed	279	1.16
Files differ	667	2.78

My goals

When we need to parse  
L<sup>A</sup>T<sub>E</sub>X

Parsers versus  
converters

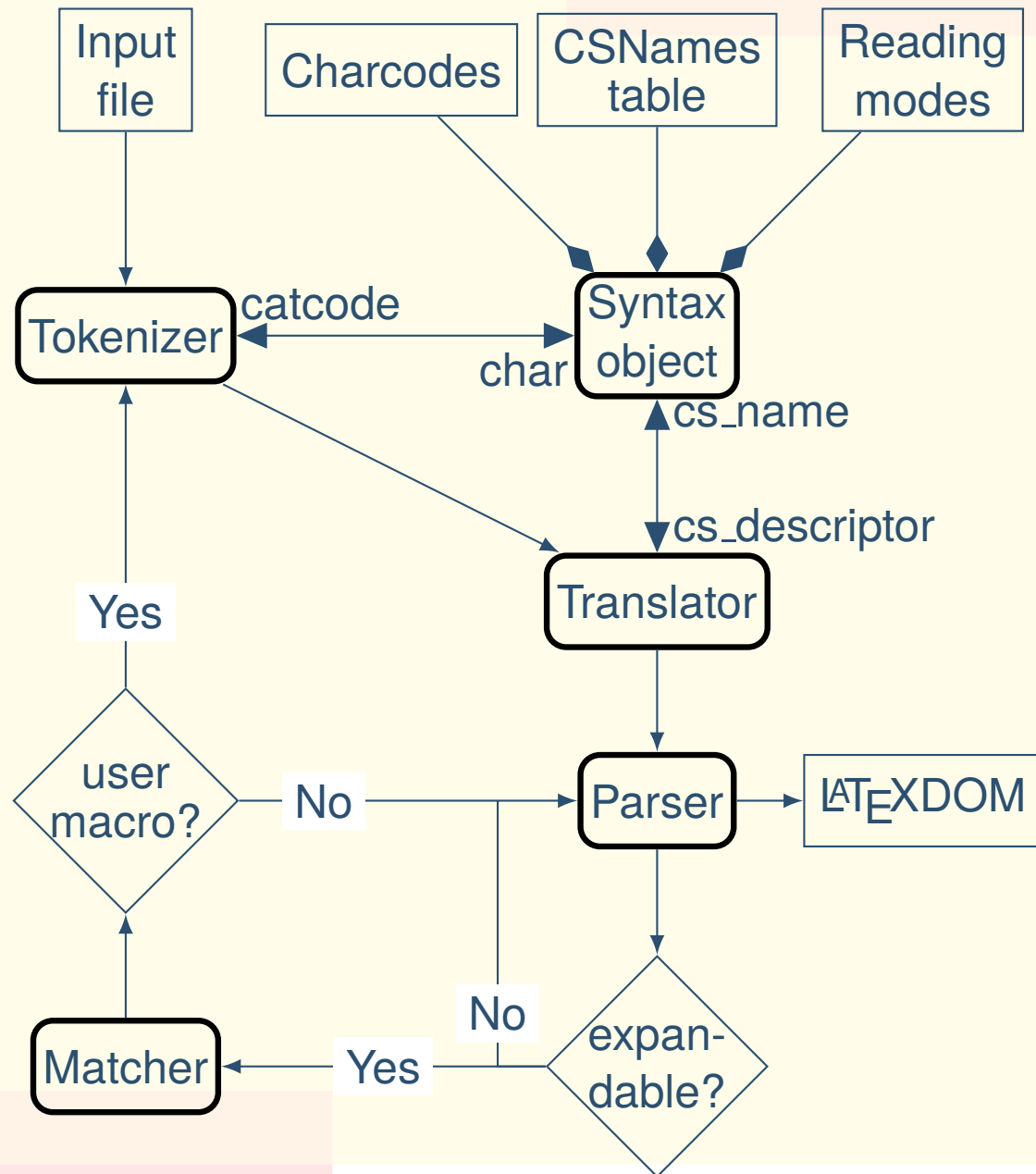
Our ongoing parser  
project

The current parsing  
statistics

Scheme of the parser

To be continued...

# Scheme of the parser



My goals  
When we need to parse  
L<sup>A</sup>T<sub>E</sub>X

Parsers versus  
converters

Our ongoing parser  
project

The current parsing  
statistics

Scheme of the parser

To be continued...

To be continued...

My goals  
When we need to parse  
 $\text{\LaTeX}$

Parsers versus  
converters

---

Our ongoing parser  
project

---

The current parsing  
statistics

Scheme of the parser

To be continued...

Thank you for your time!