

AutoBeamer 开发文档

Log Creative

版本 0.1 (2021 年 4 月 15 日)

目录

第一章 开发总览	2
第二章 词法分析器	3
2.1 词法单元	3
2.2 构建 DFA	4

第一章 开发总览

AutoBeamer 实用程序旨在转换普通的 \LaTeX 2_ϵ 讲义文件翻译为可以编译为 Beamer 幻灯片的对应文档。

该程序尚处于开发初期。决定采用 Rust 语言编写该实用程序。

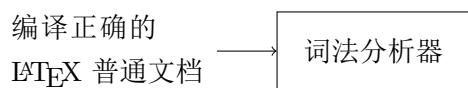


图 1.1: 目前的技术进度

第二章 词法分析器

该部分¹旨在建立对于 $\{\text{L}\text{T}_{\text{E}}\text{X } 2_{\epsilon} \text{ 命令}\} - \{\text{T}_{\text{E}}\text{X 命令}\}$ 子集的词法分析器 (lexer)。这些命令用于 AutoBeamer 识别文档分块。

2.1 词法单元

AutoBeamer 主要识别两种基本的词法单元，正则描述如表 2.1 所示：

- 命令（比如：`\title{Slide Title}`，命令参数是可选的）
- 英文词（比如：beamer）
- 其他语言的单字（比如：幻，灯，片）

表 2.1: AutoBeamer 所识别的基本词法单元

词法单元	非正式描述	正则描述
command	命令	<code>\\w+(\[.*\] \{.*\})*</code>
word	词	<code>\w+</code>
character	字	<code>[^(\ \\ \\w \\s)]</code>

此外，对于命令语法单元，AutoBeamer 需要进行扩展识别，其相关正则描述如表 2.2 所示：

- 环境开始（比如：`\begin{document}`）
- 环境结束（比如：`\end{document}`）
- 文档类（比如：`\documentclass{article}`）

¹该部分将会被用于编译原理课程的第5次作业第2题。

表 2.2: AutoBeamer 所识别的命令扩展词法单元

词法单元	非正式描述	正则描述
begin	环境开始	<code>\\begin{\w+}</code>
end	环境结束	<code>\\end{\w+}</code>
class	文档类	<code>\\documentclass([\w*\])?\{\w+\}</code>

2.2 构建 DFA

确定的有穷自动机 (Deterministic Finite Automata, DFA) 是构建词法分析器的重要依据。

基本的词法单元需要先识别命令，再识别词/字，其 DFA 如图 2.1 所示。注意，输入的文件应该是已经可以通过 L^AT_EX 编译的普通文档（不会出现 Error 异常，能输出 PDF 文档），所以 \ 后面应该就是字母，否则会报错。

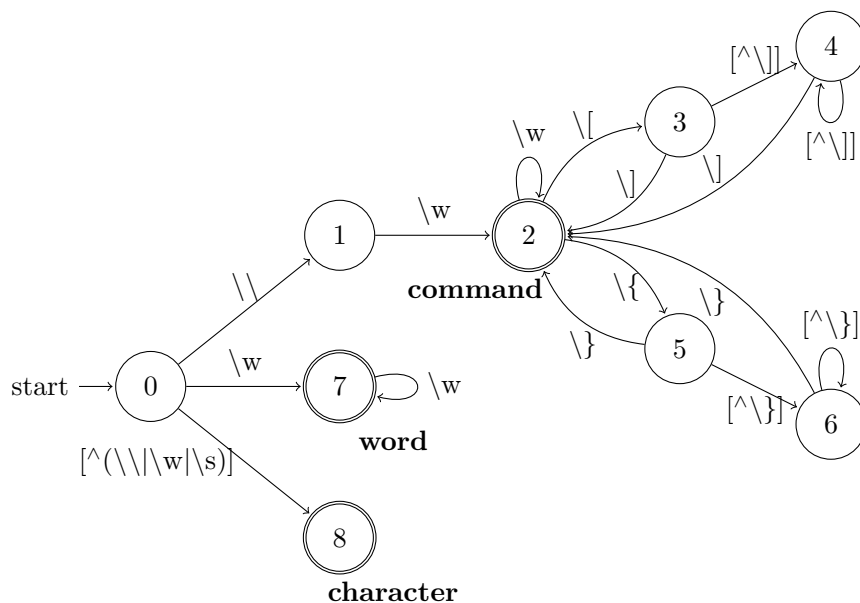


图 2.1: 基本词法单元的 DFA

对于命令扩展词法单元而言，需要有其已经是命令的先决条件，其 DFA 如图 2.2 所示。仍然是在文档是编译正确的前提下进行的，如果没

有匹配到扩展命令，将仍然作为 **command** 词法单元。

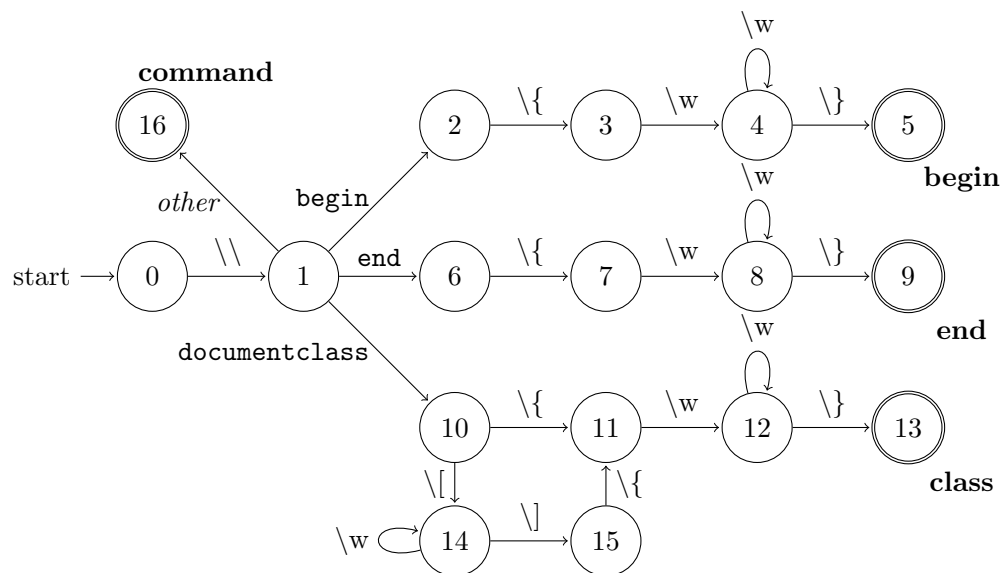


图 2.2: 命令扩展词法单元的 DFA