Report for LaTeX to Markdown Converter

Kailash Mandal

August 2024

1 Introduction

This report covers the development and implementation of a LaTeXto Markdown converter. The objective was to create a tool that parses LaTeX files and generates equivalent Markdown documents. The project involves developing both a lexer and a parser from scratch without relying on any external libraries for Markdown conversion.

2 Problem Statement

The goal of this assignment was to convert LaTeX documents to Markdown format while supporting various LaTeX features such as sections, subsection, headings, formatting (bold, italic), lists, images, tables and code blocks.

3 Features Implemented

1. Sections to Headings: Conversion of LaTeX sections to appropriate Markdown headings.
2. Text Formatting: Handling of bold and italicized text.
3. Horizontal Lines: Recognition and conversion of horizontal lines.
4. Paragraphs: Detection and proper conversion of paragraphs.
5. Code Blocks: Conversion of LaTeX code environments into Markdown code blocks.
6. Hyperlinks: Handling LaTeX hyperlinks and converting them to Markdown links.
7. Images: Processing and converting LaTeX image commands into Markdown image syntax.
8. Lists: Conversion of ordered and unordered lists from LaTeX to Markdown.

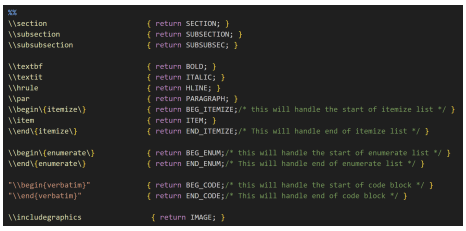

Figure 1: latex features

9. Tables: Conversion of LaTeX tables into Markdown table format.

4 Tools and Technologies

1. **Lexer:** The lexical analysis was done using Flex (latex2md.l). It scans the LaTeX input file and breaks it down into tokens.
2. **Parser:** The parsing was done using Bison (latex2md.y). The parser interprets the sequence of tokens and builds an abstract syntax tree (AST) which is then mapped to the corresponding Markdown output.
3. **Shell Script:** A run.sh script was created to automate the conversion process. It takes a LaTeX file as input and outputs the equivalent Markdown file.

5 Code Structure

- Lexer (latex2md.l): This file contains rules for recognizing LaTeX syntax and generating tokens.
- Parser (latex2md.y): This file defines the grammar of the LaTeX language and the actions to be taken to convert the parsed tokens into Markdown.
- Run Script: A shell script (run.sh) was provided to facilitate running the conversion by executing *./run.sh input.tex output.md*.

6 Challenges and Solutions

- Parsing Complex Structures: Dealing with nested environments and commands in LaTeX required careful attention to grammar rules and AST mapping.
- Handling Images and Code Blocks: Special care was taken to handle LaTeX-specific environments and convert them to their Markdown equivalents.

- Errors & warnings: encountered & handled many incompatible pointer type, shift/reduce conflicts, issues with generating AST.

7 Conclusion

This assignment provided a comprehensive exercise in compiler construction and text processing. By implementing a LaTeX to Markdown converter, we gained practical experience with tools like Flex and Bison, as well as a deeper understanding of parsing and lexical analysis.
This report, along with the code files (latex2md.l, latex2md.y, and run.sh), has been compressed into a tar file named [entry number].tar for submission on Moodle.

8 References

- https://web.mit.edu/gnu/doc/html/flex_1.html
- https://cs.wmich.edu/ yang/teach/cs485/yacc.pdf
- https://www.oreilly.com/library/view/flex-bison/9780596805418/ch01.html