ECOTE - preliminary project
Translator of a LaTeX subset to HTML

Krzysztof Rudnicki, 307585
Semester: 2023L

April 26, 2023

1 General overview and assumptions

initial task proposals (at least: assumptions, variant selection, implementation technology, scope, etc.).

My task is to create a translator of LaTeX subset to selected text format with focus on LaTeX tables

I decided to change to translator of LaTeX subset to HTML since I know LaTeX very well and HTML relatively well, I decide to translate LaTeX into HTML since HTML is easy, a little bit different than LaTeX and popular which makes this translator a practical tool.

1.1 Assumptions

- No LaTeX (%) comments in the script

- There are no extra packages in LaTeX script (provided with \usepackage keyword) besides ones distributed with LaTeX

- There are no extra classes in LaTeX script besides ones distributed with LaTeX

- There is nothing between \documentclass keyword and \begin{document} keyword

- No standard LaTeX instructions are modified in the script

- "Tables" will be represented using LaTeX *table* environment

2 Functional requirements

The goal of the project is to transform .tex file to (working) .html file if the subset of .tex file is within project scope or output error message explaining why the html could not be outputed

2.1 LaTeX subset

This project will focus almost exclusively on *table* environment
more specifically table environment containing tabular inside of it

1. \documentclass{class}: Defines what layout standard LaTeXwill use

2. \begin{document}: Ends (in our case empty) preamble

3. \end{document}: Ends LaTeX document

4. \textit{Text}: Makes the text italic.

5. \underline{Text}: Underlines the text.

6. \begin{itemize} \item Item 1 \item Item 2 \end{itemize}: Creates an unordered list with the given items.

7. \begin{enumerate} \item Item 1 \item Item 2 \end{enumerate}: Creates an ordered list with the given items.

8. \begin{center} Text \end{center}: Centers the text.

9. \begin{tabular}{|c|c|c|} \hline Column 1 & Column 2 & Column 3 \ \hline Row 1, Column 1 & Row 1, Column 2 & Row 1, Column 3 \ Row 2, Column 1 & Row 2, Column 2 & Row 2, Column 3 \ \hline \end{tabular}: Creates a table with the given columns and rows.

3 Implementation

I decided to use Python as a language in which I will implement my solution
The reasons for using python are as follow:

1. It is the easiest language among those that I know

2. I know it enough to be confident in my ability to implement this solution in python

3. I want to learn python more through this project

Negative aspects of python which is that it is very slow language do not bother me as I believe the project scope will not be big enough for this to become an issue

3.1 General architecture

3.2 Data structures

3.3 Module descriptions

3.4 Input/output description

Input is a .tex file (LaTeX file)
Outpus is an .html file
In case of errors error message will be outputed on the terminal
Input File path is entered as an argument to terminal with "-i" or "–input" flag for example:

```
python main.py −i texFile.tex
```

Output file path can be named by user by using "-o" or "–output" flag:

```
python main.py −i texFile.tex −o htmlFile.html
```

If no "-o" flag is issued the output file will have the same name as input file with changed extension to html (so in this example texFile.tex will become texFile.html)
If the path to file name consists of spaces, path name needs to be but in ""

```
python main.py −i "My␣Folder/input.tex"
```

3.5   Others

4   Functional test cases

Title	Input (LaTeX)	Output
empty file		Error! expected \documentclass at the begining of LaTeX file
Document class	\documentclass[options]{class}	Error! expected \begin{document} after document class
Extra text between document class and begin document	\documentclass[options]{class} "extra text" \begin{document}	Error! unexpected text between document class and begin document
Just document class and begin document	\documentclass[options]{class} \begin{document}	Error! no \end{document} at the end of LaTeX code

Title	Input (LaTeX)	Output
Just document class and begin/ end document	\documentclass[options]{ class} \begin{document} \end{document}	<html> </html>
Plain text inside	\documentclass[options]{ class} \begin{document} Lorem ipsum dolor sit amet. \end{document}	<html> Lorem ipsum dolor sit amet. </html>
Reduntant \ end{ document} (ignored)	\documentclass[options]{ class} \begin{document} Lorem ipsum dolor sit amet. \end{document} \end{document}	<html> Lorem ipsum dolor sit amet. </html>
LaTeX comments	\documentclass[options]{ class} \begin{document} Lorem ipsum dolor sit amet. % some comment \end{document} \end{document}	Error! LaTeX comment detected at line 3