Lex operations:

6 tags were defined using regex:
- START: html starting tags like <html>, <h1> etc.
- END: html closing tags like </html>, </h1> etc.
- NOCLOSE: html self closing tags like
, <img/> etc.
- COMMENT: html comments of the format <!-- comment -->
- DOCTYPE: starting line of html code defining the doctype of the format <!DOCTYPE ... >
- TEXT: The text in the html code

Yacc operations:

Grammar was defined as:
- doc : start doc end
  | sentence doc
  | selfclose doc
  | doctype doc
  | comment doc
  | start end
- start : START
- end : END
- sentence : TEXT
- selfclose : NOCLOSE
- doctype : DOCTYPE
- comment : COMMENT

AST:
The following information was stored in each node of the ast:
- type: whether it is a start tag, end tag, self closing tag, text etc.
- value: the name of the tag in capitals, for example, for <html>, value is HTML, for <title> it is TITLE etc.
- attr: list of list of attributes associated to the node. For text and tags which do not have any attributes, list is empty. For <img src='logo.png' width='7'>, attributes are [['src', 'logo.png'], ['width', '7']]
- children: list of children of the node

Translating the AST:
A mapping(in the form of a dictionary) is maintained for each html tag to its corresponding latex tag in the file html2latexmapping.py. The tree is traversed in a recursive manner and each html tag is mapped to its corresponding latex tag.

Programming Language:
Python

Extra things:
Handled all the Greek Letters and special characters