

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
DEPARTAMENTO ACADÊMICO DE INFORMÁTICA
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

ADSON WHOBBERT DA LUZ
LUIZ HENRIQUE KULTZAK

**TRABALHO PRÁTICO 2: CONSTRUÇÃO DE UM TRADUTOR DE
LATEX PARA HTML**

PONTA GROSSA

2016

ADSON WHOBBERT DA LUZ
LUIZ HENRIQUE KULTZAK

**TRABALHO PRÁTICO 2: CONSTRUÇÃO DE UM TRADUTOR DE
LATEX PARA HTML**

Trabalho apresentado à disciplina de
Compiladores, como requisito parcial para
obtenção de nota no segundo bimestre.

Professor: Gleifer Vaz Alves

PONTA GROSSA
2016

SUMÁRIO

2.1 LATEX.....	4
2.2 HTML.....	5
2.3 FLEX E BISON.....	5
3.1 GRAMÁTICA.....	6
3.2 ANALISADOR LÉXICO.....	7
3.3 ANALISADOR SINTÁTICO.....	7
3.4 Testes.....	8

1 INTRODUÇÃO

O presente trabalho consiste na implementação de um analisador léxico e sintático para a tradução de um arquivo contendo texto no formato LATEX para um arquivo no formato html. O arquivo html deve manter as estruturas presentes no arquivo LATEX tais como parágrafos, listas enumeradas, seções em negrito, etc.

2 FERRAMENTAS

Esta seção está destinada a explicação das ferramentas que são utilizadas no trabalho. A ferramenta LaTeX para criação de documentos PDF formatados. A ferramenta HTML para geração de páginas WEB. As ferramentas Flex e Bison para construção de um analisador léxico e sintático.

1 LATEX

LaTeX¹ é um sistema de preparação de documentos para composição de alta qualidade. Sua maior utilidade é para a geração de documentos técnicos ou científicos de médio a grande porte, todavia pode também ser utilizado para quase qualquer forma de publicação.

LaTeX baseia-se na ideia de que é melhor deixar o design do documento para os projetistas de documentos e permitir que os autores continuem escrevendo documentos. Então, no LaTeX você entraria este documento como:

```
\documentclass{article}
\title{Título do documento}
\author{Fulano da Silva}
\date{Setembro 2016}
\begin{document}
  \maketitle
  Corpo do documento
\end{document}
```

Traduzindo seu funcionamento:

- Este documento é um artigo.
- Seu título é “Título do documento”.
- Seu autor é Fulano da Silva.
- Foi escrito em setembro de 2016.
- O documento é composto por um título seguido pelo texto Corpo do documento

¹ <http://www.latex-project.org/about/>

2 HTML

HTML² é a linguagem de marcação padrão para criar páginas da Web. Arquivos no formato HTML são executados por navegadores de internet que utilizam as marcações de tags do documento .html para definir como serão exibidos os elementos explicitados.

As tags em HTML são nomes de elementos rodeados por colchetes angulares:

`<nome da tag>conteúdo aqui...</nome da tag>`

Um exemplo de elemento utilizado na linguagem HTML é a tag `<p>`, que define um parágrafo a ser exibido pelo navegador. Exemplo:

`<p>` O conteúdo inserido dentro deste contexto será considerado um parágrafo pelo interpretador `</p>`.

3 FLEX E BISON

Flex e Bison são ferramentas para a construção de programas que manipulam entradas estruturadas. Eles eram originalmente ferramentas para construir compiladores, mas provaram ser úteis em muitas outras áreas.

A ferramenta Flex atua como um analisador léxico de entradas. A tarefa principal do analisador léxico é ler os caracteres da entrada do programa fonte, agrupá-los em lexemas e produzir como saída uma sequência de tokens para cada lexema no programa fonte. O fluxo de tokens é enviado ao Analisador Sintático para que a análise seja efetuada.

² http://www.w3schools.com/html/html_intro.asp

3 IMPLEMENTAÇÃO

4 GRAMÁTICA

A notação Backus-Naur (mais conhecida como BNF ou Backus-Naur Form) é uma forma matemática formal de descrever uma linguagem, desenvolvida por John Backus (e possivelmente Peter Naur também) para descrever a sintaxe da linguagem de programação Algol 60.

Para a visualização da gramática utilizada para solucionar o problema proposto foi utilizada a notação BNF.

Segue a definição da gramática desenvolvida para a solução do problema proposto:

`entrada ::= tipoDocumento INICIODOCUMENTO corpo FIMDOCUMENTO`

`tipoDocumento ::= classe
 | tipoDocumento autor
 | tipoDocumento titulo`

`autor ::= AUTOR ABRECHAVES texto FECHACHAVES`

`classe ::= CLASEDODOCUMENTO ABRECHAVES texto FECHACHAVES`

`titulo ::= TITULODOCUMENTO ABRECHAVES texto FECHACHAVES`

`corpo ::= corpo opcao
 | opcao`

`opcao ::= textoFormatado
 | lista
 | paragrafo
 | capitulo
 | secao
 | subsecao
 | comentario`

`comentario ::= COMENTARIO textoFormatado`

`lista ::= INICIOLISTAE itemLista FIMLISTAE
 | INICIOLISTAI itemLista FIMLISTAI`

`itemLista ::= ITEM textoFormatado itemLista`

|

```

textoFormatado::= NEGRITO ABRECHAVES textoFormatado FECHACHAVES
| ITALICO ABRECHAVES textoFormatado FECHACHAVES
| SUBLINHADO ABRECHAVES textoFormatado FECHACHAVES
| ABRECHAVES NEGRITO textoFormatado FECHACHAVES
| ABRECHAVES ITALICO textoFormatado FECHACHAVES
| ABRECHAVES SUBLINHADO textoFormatado FECHACHAVES
| texto

```

```

paragrafo::= PARAGRAFO ABRECHAVES textoFormatado FECHACHAVES

```

```

capitulo::= CHAPTER ABRECHAVES textoFormatado FECHACHAVES

```

```

secao::= SECAO ABRECHAVES textoFormatado FECHACHAVES

```

```

subsecao::= SUBSECAO ABRECHAVES textoFormatado FECHACHAVES

```

```

texto::= texto PALAVRA
| PALAVRA

```

5 ANALISADOR LÉXICO

O analisador léxico foi implementado na ferramenta Flex e nela foram definidos os seguintes *tokens*:

```

BARRAINVERTIDA,      ABRECHAVES,      FECHACHAVES,
CLASEDODOCUMENTO,   TITULODOCUMENTO, PALAVRA, INICIOLISTAE,
FIMLISTAE,          INICIOLISTAI,    FIMLISTAI,  ITEM,  INICIODOCUMENTO,
FIMDOCUMENTO, PARAGRAFO, CHAPTER, SECAO, SUBSECAO, NEGRITO,
ITALICO, SUBLINHADO, AUTOR, COMENTARIO, PALAVRA.

```

No analisador léxico foram utilizadas estruturas para controlar a criação de capítulos, seções e subseções. A variável *insec* foi utilizada para controlar o fluxo de seções. A variável *subsec* para controlar o fluxo de subseções e por fim a variável *inchap* para controlar o fluxo de capítulos do documento.

6 ANALISADOR SINTÁTICO

A partir da gramática descrita na seção 3.1, foi gerado o código Bison responsável por traduzir cada *tag* do documento LaTeX para *tag* HTML. Fazendo uso da função *fprintf* para imprimir no arquivo de saída.

Um exemplo desse procedimento pode ser visualizado a seguir:

Gramática:

```
texto ::= texto PALAVRA
        | PALAVRA
```

Código Bison:

```
texto: texto PALAVRA {fprintf(novoHTML," %s",$2);}
      | PALAVRA {fprintf(novoHTML," %s",$1);}
      ;
```

7 TESTES

Foram criados 3 arquivos no formato LaTeX para teste das funcionalidades implementadas. Bem como um arquivo *makefile* responsável por organizar a compilação dos arquivos Flex e Bison e execução dos exemplos.

Segue texto do arquivo *makefile*:

```
saida: flex.l bison.y
      bison bison.y -d
      flex flex.l
      gcc -o trabalho2.exe bison.tab.c lex.yy.c -lfl
      ./trabalho2.exe saida.html < entrada.tex > log.txt
      ./trabalho2.exe saida2.html < entrada2.tex > log.txt
      ./trabalho2.exe saida3.html < entrada3.tex > log.txt
remove:
rm bison.tab.c saida1.html saida2.html saida3.html trabalho2.exe lex.yy.c
log.txt bison.tab.h
```

No exemplo a seguir é possível observar a tradução LaTeX para HTML:

Trecho do documento LaTeX:

`\bf{Latex}` (normalmente formatado como `\bf{LaTeX}`) e um conjunto de macros para o programa de diagramacao de textos `\textit{TeX}`, utilizado amplamente na producao de textos matematicos e cientificos, devido a sua alta qualidade tipografica. Entretanto, tambem e utilizado para producao de cartas pessoais, artigos e livros sobre assuntos diversos.

Trecho equivalente do documento HTML:

```
<b> Latex</b>
(normalmente formatado como<b> LaTeX</b>
) e um conjunto de macros para o programa de diagramacao de textos<i>
TeX</i>
, utilizado amplamente na producao de textos matematicos e cientificos,
devido a sua alta qualidade tipografica. Entretanto, tambem e utilizado para
```

producao de cartas pessoais, artigos e livros sobre assuntos diversos.
2
Utilizacao</br>

4 CONCLUSÃO

A junção das ferramentas Flex e Bison possibilitam a implementação de inúmeras funcionalidades. Foi possível observar durante a execução deste trabalho uma aplicação direta proveniente do uso dessas ferramentas.

REFERÊNCIAS

http://www.w3schools.com/html/html_intro.asp

<http://www.latex-project.org/about/>