

Universidade do Minho

DEPARTAMENTO DE INFORMÁTICA

Processamento de Linguagens Trabalho Prático nº 2

Gonçalo Almeida (A84610)

Emanuel Rodrigues (A84776)

Lázaro Pinheiro (A86788)

28 de Junho de 2020

Conteúdo

1	Introdução	3
2	Objetivos do trabalho prático	4
3	Notas iniciais	5
4	Implementação 4.1 Flex <	6 6 6 7 7 9 9
5	Resultado Final	11
6	Conclusão	13
A	Flex	14
В	Yacc	15

Introdução

Este projeto foi realizado no âmbito da unidade curricular Processamento de Linguagens e tem como objetivo construir um processador (flex, yacc) que reconheça e valide cadernos de anotação gerando um site HTML. Para isto foi utilizada a ferramenta Flex de modo a processar o texto de entrada, o Yacc para escrever a gramática que define a DSL de caderno de anotações e gerar o site HTML e a biblioteca Glib para o armazenamento dos conteúdos.

Objetivos do trabalho prático

- Aumentar a experiência de uso do ambiente Linux , da linguagem imperativa C (para codificação das estruturas de dados e respetivos algoritmos de manipulação), e de algumas ferramentas de apoio à programação;
- Rever e aumentar a capacidade de escrever gramáticas independentes de contexto (GIC) , que satisfaçam a condição LR(), para criar Linguagens de Domínio Específico (DSL);
- Desenvolver processadores de linguagens segundo o método da gramática tradutora (GT) ;
- utilizar geradores de compiladores como o par flex/yacc.

Notas iniciais

Dado que o nosso grupo é o 67, usando a formula que se encontra no enunciado (exe = (NGr%6) + 1), foi nos atribuído o trabalho 2, DSL para Cadernos de Anotações em HD.

Implementação

4.1 Flex

Com a ajuda da biblioteca Glib definimos uma string para armazenar o conteúdo processado de um documento:

GString* conteudo = g_string_new(NULL);

4.1.1 Filtros de Texto Flex

De modo a processar o ficheiro de entrada desenvolvemos um ficheiro Flex que faz o parse de todos os campos pedidos para o Yacc.

Conceito

Para filtrar o conceito do documento decidimos capturar o texto que vem a seguir a ===\\: até encontrar um \n.

Título

Para capturarmos o título do documento, guardamos todo o texto que aparece a seguir a @tit $\:$ até a um $\n.$

Subtítulo

O subtítulo é obtido através da expressão <TEXTO>#\ .+[\n]2

Conteúdo

Para se filtrar o conteúdo usamos a expressão <CONT>.+/\n que vai dar append no array conteúdo do texto encontrado, quando é encontrado um \n no inicio da linha ou um novo subtítulo, o conteúdo e passado ao yacc e o array e apagado.

Triplos

Quando se encontra um @triplos\: vamos entrar na condição de contexto do sujeito do triplo, dentro dela guardamos o sujeito através de <SUJ>\:.+/\n e de seguida entrámos na condição de contexto da relação, nela retiramos o texto da relação e entramos na condição do objeto ou então vamos para a condição da imagem se se encontrar um \:img/\, guardamos o nome da imagem

e voltamos para a relação se for encontrado um ; ou para o sujeito para um \backslash . No objeto procedemos de maneira semelhante a imagem.

4.2 Yacc

4.2.1 Gramática

A abordagem tomada para o desenvolvimento da gramática que melhor representa lingua- gem a reconhecer foi uma top-down.

Cada especificação de um caderno consiste numa lista de símbolos não terminais 'par'.



Cada especificação de um 'par' encontra-se essencialmente dividida em duas partes, uma em que se encontra o 'documento' (documento) e outra em que se encontra o 'triplos' (triplos do documento).

```
par : documento triplos
;
```

Cada especificação de um 'documento' deriva nos símbolos terminais 'CONCEITO' (respetivo ao conceito em causa) e 'TITULO' (respetivo ao título do conceito), e no símbolo não terminal 'topicos' onde se encontram os tópicos de cada documento.

```
documento : CONCEITO TITULO topicos
```

Cada especificação de 'topicos' consiste numa lista de símbolos terminais 'SUBTITULO' (representantes dos subtítulos de cada tópico) e de símbolos não terminais 'texto' (representantes do texto respetivo a cada subtítulo).

```
topicos : topicos SUBTITULO texto
```

Cada especificação de 'texto' consiste numa lista de símbolos não terminais 'CONTEUDO' (representantes de cada parágrafo do respetivo texto).



Cada especificação de 'triplos' consiste numa lista de símbolos terminais 'SUJEITO' (representantes dos sujeitos dos triplos) e de símbolos não terminais 'relacoes' (representantes das relações respetivas a cada sujeito).

```
triplos : triplos SUJEITO relacoes
```

Cada especificação de 'relacoes' consiste numa lista de símbolos terminais 'RELACAO' (representantes das relações dos sujeitos) e de símbolos não terminais 'objectos' (representantes das objetos respetivos a cada relação).



Cada especificação de 'objectos' consiste numa lista de símbolos terminais 'OBJECTO' (representantes de cada objeto da respetiva relação) ou de símbolos terminais 'IMAGEM' (representantes de cada imagem do respetivo sujeito).



4.2.2 HTML

Para uma melhor navegação nos ficheiros gerados, foi criado um índex (index.html) que contêm as referências para todos os conceitos encontrados no ficheiro de entrada.

O HTML gerado para cada sujeito/objeto apenas consiste no seu título, imagens e botão para regressar ao índex gerado.

O HTML gerado para cada conceito (sendo também sujeito/objeto ou não) consiste nos mesmos elementos referidos anteriormente mais todos os seus tópicos e a listagem de todos os seus triplos, estes contendo as referências para os sujeitos e objetos.

E de notar que as imagens são todas guardadas em memória e apenas adicionadas ao ficheiro no final do parse do ficheiro de entrada de modo a que estas não fiquem espalhadas, mas sim agrupadas numa classe com propriedades independentes.

4.2.3 Estruturas de dados

Como foi referido acima, as imagens são guardadas em memória. Para tal foram definidas algumas estruturas de dados. Algumas destas estruturas são referentes à biblioteca Glib que facilitou o seu manipulamento.

A estrutura 'Imagens' armazena, para cada sujeito, as respetivas imagens e o número destas.

```
typedef struct imagens {
   char* sujeito;
   int numImgs;
   char* imgs[NUM_MAX_IMAGES];
} *Imagens;
```

A estrutura 'buff' é um GArray que armazena todas as estruturas 'Imagens' referidas anteriormente.

GArray *buff;

A estrutura 'imgs' é um array que armazena as imagens do sujeito a ser tratado momentariamente. O inteiro 'imageCount' é referente ao número de imagens nesta estrutura.

```
char* imgs[NUM_MAX_IMAGES];
int imageCount = 0;
```

Resultado Final

Hitler

Adolf Hitler

Vida pessoa

Hiber quarria passar a imagem de um homem cellostário que não tinha vida doméstica, dedicando-se inteliamente a sua missão política e a sua nação. Ele conheceu sua principal amante, Eva Braun, em 1929, e se casou com ela em abril de 1945. Em setembro de 1931, sua meia-sobrinha, Gell

Visões religiosa

Htler nasceu de uma mãe que era católica praticante e um pai anticlerical; após deixar a sua casa, Hiller praticamente nunca mais frequentou missas ou recebeu sacramentos. Albert Speer afirmou que Htler fez pronunciamentos duros contra a Igreja para seus associados políticos e apesar de não ter abandonado a 6f., nunca Igou multo para eia. Htler dizia que se os fiéis abandonassem a Igreja eles iriam se virar ao misticismo, o que ele considerava um retrocesso. De acordo com Speer, o Fibrer acreditava que a religião paponesa ou o islamismo seriam religiões melhores para os alemães d

O historiador John S. Conway afirmou que Hilder se opunha fundamentalmente as igrejas cristãs. De acordo com Alan Bullods, Höber não acreditava em Deus, era anticierical e não acreditava muito na 'ética cristã' por que ela ia de encontro a sua crença de "sobrevivência do mais apto". Ele

Hitter via a ligniqui como uma importante influência política conservadora na sociedade, e adotou uma estratégia para lidar com eles que "se encaixava com os seus propóstos políticos imediatos". Em público, Hitter exaltava a herança cristá alemá e a sua cultura, professando a sua crença no "Jesus Ariano", que havia lutado contra os judeus. Muitas das suas retóricas pri-cristás não eram as mesmas que ele fazia em privado, descrevendo o cristanismo como o "abourdo" e fundado em mentiras.

De acordo com um relatório da Agência de Serviços Secretos dos Estados Unidos, intituíado "The Nazi Master Plan" ("O Plano mestre Nazista"), Hitler planeave destruir a influência da igreja cristã dentro do Reich. O seu plano final seria a eliminação total do cristianismo. De acordo com Bullock, Hitler Iria executar tal plano apris a conclusão da guerra na Europa.

Speer secreveu que Hitler também tinha visões negativas a respeito das tendêncies ao misticismo de Heinrich Himmler e Alfred Rosenberg. O führer não gostava da tentativa de Himmler de tentar mitilicar a SS. Hitler era mais pragmático e as suas ambições centravam-se em preocupações mai

Pessoas próximas a Hitler afirmavam que ele era teísta. Frequentemente, ele mencionava que a "Providência" o protegia e dizia estar em uma missão divina na terra para eliminar os judeus e reerguer o povo alemão.

Culto à personalidad

Sob seu governo, Hitler foi objeto de um extenso culto de personalidade. Seus retratos foram pendurados em locais públicos em toda a Alemanha e a saudação Heil Hitler ("Salve Hitler") tomou-se obrigatória mesmo entre civis. Este culto era intrinsecamente ligado ao modelo ditatorial adotado pel Alemanha, com doutrinadores nazistas justificando-o com base no chamado principio da liderança (em alemão: Führerprincipi), segundo o qual toda a autoridade era investida em Hitler, que estava necessariamente certo e deveria ser inquestionavelmente obedecido tanto pelà hierarquia nazista como poliço pos alemanãos, em inalicade desmológica e capa.

Para assegurar a máxima difusão de seus ideais, foi instituído logo após a tomada do poder pelos nazistas em 1933 um Ministério da Propaganda, que até 1945 manteve total controle sobre as produções culturais no país. Presidido por Josef Geebbels, o Ministério teve Hiller como um dos principais focos de proaganda, plorificação-como um lidor herorico a infalíve e contribuindo fortemente para seu culto de personalidade. Hilter a paraceus ele memo em mutios de finnes, performando de forma culádocamente corregoriada. As Reunides de Numemberg aram um dos principais eventos para promoção da imagen de Hilter, e foram acces mos insella, entribunido-be uma concatidado politica negativa.

Conceitos

- Salazar
- Jong-ur

```
Hillier

Busser

Dusder

Publico

Bibrer

Millier

Barcher

Antifolis

Nome

Antifolister

lord, place

Browner, antifolister

lord, place

Browner, antifolister

lord, place

Browner, antifolister

Browner, antifolister

Autoriales

Antifolis

Person de Antifolis

Person de Antifolis

Browner, and Inco

particia, publico

Browner, and Inco

particia, publico

Person de Antifolister

Browner, and Inco

particia, publico

Person de Antifolister

Browner, and Inco

particia, publico

Person de Antifolister

Browner, and Antifolister

Examina

Antifolis

Nation

Antifolis

Person de Antifolister

Examina

Exami
```

Imagens

Conclusão

A realização deste trabalho prático permitiu compreender a grande utilidade da ferramenta Flex no tratamento de informação e o Yacc permite a especificação de uma hierarquia complexa de relacionamentos entre conceitos. Com recurso a esta é gerado um ficheiro html que resulta numa página web. Esta página web permite o link com outras páginas web , permitindo assim a consulta de conceitos, de triplos e também a visualização das imagens.

Numa primeira abordagem o grupo não percebeu o que estava a ser requerido no enunciado, contudo a sessão de dúvidas facilitou bastante este processo. Em suma, obtivemos uma melhor compreensão e prática relativamente às expressões regulares bem como consolidamos o conhecimento sobre gramáticas adquirido no decorrer das aulas da U.C. . É necessário frisar que com este trabalho aprofundamos o nosso conhecimento teórico e prático relativamente à linguagem de programação html, considerando assim que os requisitos propostos foram cumpridos.

Apêndice A

Flex

```
#include<glib.h>
#include "y.tab.h"
void yyerror(char* s);
%option noyywrap yylineno
%x TEXTO CONT SUJ REL OBJ IMG
                                                                      GString* conteudo = g_string_new(NULL);
===\ \:.+/\n
                                                                      { yylval.string = strdup(yytext+5); return CONCEITO; }
                                                                      { BEGIN TEXTO; yylval.string = strdup(yytext+6); return TITULO; }
@tit\:\ .+/\n
<TEXTO>@triplos\:
<TEXTO>#\ .+[\n]{2}
<TEXTO>\n
                                                                       { BEGIN SUJ; } { BEGIN CONT; yytext[strlen(yytext)-2] = '\0'; yylval.string = strdup(yytext+2); return SUBTITULO; }
<CONT>.+/\n
<CONT>^\n
<CONT>\n/(#|@)
                                                                       { g_string_append(conteudo, yytext); }
{ yylval.string = strdup(conteudo->str); g_string_erase(conteudo, 0, -1); return CONTEUDO; }
{ BEGIN TEXTO; yylval.string = strdup(conteudo->str); g_string_erase(conteudo, 0, -1); return CONTEUDO; }
{ g_string_append(conteudo," "); }
<CONT>\n
<SUJ>\n/===\ \:
<SUJ>\:.+/\n
<SUJ>\n
                                                                       { BEGIN 0; } { BEGIN REL; yylval.string = strdup(yytext+1); return SUJEITO; }
                                                                       { BEGIN IMG; }
{ BEGIN OBJ; yylval.string = strdup(yytext); return RELACAO; }
{ BEGIN OBJ; yylval.string = strdup(yytext+1); return RELACAO; }
.
<REL>\:img/\
<REL>a/\
<REL>\:[^]*/\
<REL>[\n;,]
<OBJ>\.
<OBJ>;
<OBJ>\:[^,;,]*/(\ |,|;\\.)
<OBJ>\".*/\"
<OBJ>[,\n\"]
                                                                       { yylval.string = strdup(yytext+1); return OBJECTO; } { yylval.string = strdup(yytext+1); return OBJECTO; }
                                                                       { BEGIN SUJ; }
{ BEGIN REL; }
{ yylval.string = strdup(yytext+1); return IMAGEM; }
<IMG>\.
<IMG>;
<IMG>;
<IMG>\".*/\"
<IMG>[,\n\"]
.|\n
%%
```

Apêndice B

Yacc

```
#define _GNU_SOURCE
#define FALSE 0
#define TRUE 1
#define NUM_MAX_IMAGES 5
    #include <stdio.h>
   #include <stdlib.h>
#include <stdlib.h>
#include <string.h>
#include <math.h>
#include <unistd.h>
#include <fcntl.h>
#include <fcntl.h>
#include <glib.h>
    int yylex();
void yyerror(char*);
    void fileGenerator(char*, int);
void addText(char*, char*, char*);
void generateTriplo (char*);
void addTriplos(char*, char*);
char* formatName (char*);
void writeInIndexHtml(char*, char*);
    void addImages (char*);
    char* indexPath = "base/index.html":
    typedef struct imagens {
  char* sujeito;
  int numImgs;
  char* imgs[NUM_MAX_IMAGES];
    char* imgs[NUM_MAX_IMAGES];
    int imageCount = 0;
%union{
    char* string;
% token CONCEITO TITULO SUBTITULO CONTEUDO SUJEITO RELACAO OBJECTO IMAGEM
% type <string> CONCEITO TITULO SUBTITULO CONTEUDO SUJEITO RELACAO OBJECTO IMAGEM topicos texto triplos relacoes objectos documento
caderno : caderno par
par : documento triplos
                                                       { addTriplos($1, $2); }
documento : CONCEITO TITULO topicos
                                                        { fileGenerator($1, TRUE); addText($1, $2, $3); }//asprintf(&$$, "%s", $1); }
                                                        { asprintf(&$$, "%s\t\t\t\t\A3>%s</h3>\n%s", $1, $2,$3); } { asprintf(&$$, "");}
topicos : topicos SUBTITULO texto
texto : texto CONTEUDO
                                                        { asprintf(&$$, "%s\t\t\t%s\n\n", $1, $2);} { asprintf(&$$, "");}
triplos : triplos SUJEITO relacoes
                                                        relacoes : relacoes RELACAO objectos
                                                        | objectos IMAGEM
```

```
void fileGenerator(char* name, int isConcept){
     if(name){
           char *command, *path;
char * nameFormated = formatName(name);
asprintf(&path, "base/%s/%s.html", nameFormated, nameFormated);
           if (access(path, F_0K) == -1){
    asprintf(&command,"cd base;mkdir %s;cd %s;touch %s.html", nameFormated, nameFormated, nameFormated);
    system(command);
                free(command);
                FILE* file = fopen(path,"u");

fprintf(file, "<!DOCTYPE html>\n\t<\html lang=\"pt-pt\">\n\t<\html lang=\"pt-pt\">\n\t<\title>%s</title>\n\t\t<\meta charset=\"utf-8\">\n\t\t<\meta name=\"viewport\" content=\"width=device-width, fclose(file);
          if(isConcept){
   writeInIndexHtml(path+5, name);
          }
void addText(char* conceito, char* subtitulo, char* texto){
  if(conceito && subtitulo && texto){
           char* path;
           free(path);
    }
}
// Escreve os triplos
void addTriplos (char* conceito, char* texto){
   if(conceito && texto){
           char* path;
asprintf(&path,"base/%s/%s.html",conceito, conceito);
           asprint(specific property parts) as repen(parts, "a");

fprintf(file, "\t\t<\div class=\"triplos\">\n\t\t\t<\div>\n", texto);
          fclose(file);
free(path);
void generateTriplo (char* name) {
     if (name[strlen(name)-1] == ' ') name[strlen(name)-1] = '\0';
     char* path;
asprintf(&path, "base/%s", nameFormated);
     fileGenerator(name, FALSE):
     free(path);
char* formatName (char* name) {
     char* nameFormated;
     asprintf(&nameFormated, "%s", name);
     for (int i = 0; nameFormated[i] != '\0'; i++){
    if (nameFormated[i] == ' ') nameFormated[i] = '_';
     3
     return nameFormated;
void addImages (char* sujeito){
     if (sujeito[strlen(sujeito)-1] == ' ') sujeito[strlen(sujeito)-1] = '\0';
     if(imageCount > 0){
  int found = 0;
           Imagens imagens;
          for(int i = 0 ; i < buff->len && !found ; i++){
   imagens = g_array_index(buff, Imagens, i);
   if(strcmp(imagens->sujeito, sujeito) == 0){
      found = 1;
      for (int i = imagens->numImgs; i < imagens->numImgs + imageCount && i < NUM_MAX_IMAGES; i++){
            asprintf(imagens->imgs+i, "\t\t\t<img class=\"imagem\" src=\"%s\" width=\"100\" height=\"100\">", imgs[i]);
      }
}
                     imagens->numImgs += imageCount;
               }
          if(!found){
   imagens = (Imagens)malloc(sizeof(struct imagens));
   asprintf(&imagens->sujeito, "%s", sujeito);
   imagens=>numlings = imageCount;
   for (int i = 0; i < imageCount; i++){
        asprintf(imagens->imgs+i, "\t\t\cing class=\"imagem\" src=\"%s\" width=\"100\" height=\"100\">", imgs[i]);
   }
                g_array_append_val(buff, imagens);
          imageCount = 0;
```

```
void finalizeFiles(){
       char *command; asprintf(&command, "cd base; ls -d */ | sed 's/.$//' > output ");
       system(command);
       free(command);
      FILE* f = fopen("base/output", "r");
char directory[128];
char* path;
char* name;
Imagens imagens;
      if(!f){
      if(!?){
   perror("Ocorreu um erro!\n");
}else{
   while (fgets(directory, 128, f)){
      name = strtok (directory, "\n");
      asprintf(&path, "base/%s/%s.html", name, name);
      FILE* fp = fopen(path, "a");
                  free(imagens->sujeito);
                         flag = 0;
                   ffprintf(fp, "\n\t</body>\n</html>");
fclose(fp);
free(path);
             fclose(f);
            system("cd base; rm -f output");
/*
 * actions in index.html
 */
void generateIndexHtml(){
    system("mkdir base; cd base; touch index.html");
    FILE* file = fopen(indexPath, "w");
    fyrinf(file, "<!DOCTYPE html>\n\t<html lang=\"pt-pt\">\n\t<head>\n\t\t<title>index</title>\n\t\t<meta charset=\"utf-8\">\n\t\t<style>\n\t\t\tdiv.cabecalho {\n\t\t\t\text-align: center;\
    fclose(file);
\label{local_point} \mbox{\tt void writeInIndexHtml(char* path, char* name)} \{
       if(path && name){
             asprintf(&command, "cd base;grep -w -c \"%s\" index.html > output", name); system(command); free(command);
             FILE* f = fopen("base/output", "r");
char buffer[2];
             fgets(buffer, sizeof(buffer), f);
             if(buffer[0] == '0'){
   FILE* file = fopen(indexPath, "a");
   fprintf(file, "'\t\t\<a href='%s'>%s</a>\n", path, name);
   fclose(file);
           system("cd base; rm -f output");
     }
}
void finalizeIndexHtml(){
   FILE* file = fopen(indexPath, "a");
   fprintf(file, "\t\t</div>\n\t</body>\n</html>");
   fclose(file);
int main(int argc, char* argv[]){
   extern FILE *yyin;
   if(argc > 1) ( yyin = fopen(argv[i],"r"); }else{perror("Argumentos Insuficientes!");}
   buff = g_array_new(FALSE, FALSE, sizeof(Imagens));
   generateIndexHtml();
inalizeIndexHtml();
finalizeFiles();
g_array_free (buff, TRUE);
return 0;
}
       yyparse();
finalizeIndexHtml();
void yyerror(char* s){
   extern int yylineno;
   extern char* yytext;
   fprintf(stderr,"linha %d: %s (%s)\n",yylineno,s, yytext);
```

}