

Cvičení FJP

Program pro lexikální analýzu FLEX

Martin Lipinský
A09N0032K

Obsah

1	Zadání	2
2	Postup řešení	2
3	Příklady	2
3.1	Vyhledání emailových address	2
3.2	Nahrazení zavináče spojením at	2
3.3	Odstranění komentářů	3
4	Závěr	3

1 Zadání

Zadáním bylo vyzkoušet si něco smysluplného s programem lex (linuxový klon FLEX).

2 Postup řešení

Pomocí na Internetu hojně dostupné literatury a materiálů ke cvičení jsem se pokusil proniknout alespoň do základů tohoto mocného nástroje. První prográmk je výsledkem modifikace nalezeného programku, u dalších dvou jde již o plně mojí vlastní tvorbu, byť velmi jednoduchou.

3 Příklady

3.1 Vyhledání emailových address

Následující program vyhledá v zadaném textu emailové adresy. Regulární výraz byl zvolen tak, aby co nejvěrněji rozpoznával opravdový formát emailové adresy.

```
%{
#include<stdio.h>
#include<string.h>
%}

ZNAKY [_a-zA-Z0-9\~]
NEZNAKY [^a-zA-Z0-9]

%%

{ZNAKY}+(\.{ZNAKY}+)*@{ZNAKY}+(\.{ZNAKY}+)*(\.{ZNAKY}{2,4})/{NEZNAKY} {
printf("%s\n",yytext);
}

.

\n

%%
main (int argc, char **argv)
{
    yylex();
}
```

3.2 Nahrazení zavináče spojením at

Vzhledem k rozšířenému zneužití emailových address pro distribuci spamu a jejich častým strojovým vykrádání z WEBu je někdy používáno nahrazení znaku @ v emailových adresách slovním spojením at. Následující příklad to provede v zadaném textu automaticky.

```
%{
#include<stdio.h>
%}

ZNAKY [_a-zA-Z0-9\~]
```

```

%%
{ZNAKY}\@{ZNAKY} {printf("%c at ",yytext[0]); printf("%c",yytext[2]);};

%%
main (int argc, char **argv)
{
    yylex();
}

```

3.3 Odstranění komentářů

V posledním případě jsem se pokusil využít podpor stavů konečného automatu programu FLEX, a pomocí dvou stavů odstraňovat komentáře ze shell skriptu. Program hledá jakýkoli výskyt znaku # a vynechává vše co po něm následuje až do konce řádku. Jedinou výjimkou je kombinace #! uvozující v shell skriptech čím má být skript interpretován.

```

%{
#include<stdio.h>
%}

%x comment

%%

^"#!" BEGIN(INITIAL); ECHO;
^"#" BEGIN(comment);
" #" BEGIN(comment);
<comment>\n BEGIN(INITIAL);
<comment>(.\n)

%%
main (int argc, char **argv)
{
    yylex();
}

```

4 Závěr

Při cvičení jsem si vyzkoušel základní možnosti FLEXu. Zkusil jsem si jak složitější regulární výraz, tak jednoduchoučký dvojstavový konečný automat v příkladu s odstraňováním komentářů ze shell skriptů. Cvičení odhalilo pro mě dosud neznámý potencionálně velmi užitečný software, o kterém jsem, ač uživatel Linuxu neměl ani potuchy. V tomto ohledu více než splnilo svůj účel.