McMaster University
Department of Computing and Software
Dr. W. Kahl

SFWR ENG 3E03
Exercise Sheet 3
Solution Hints

Design and Selection of Programming Languages

27 September 2006

Simplified FunnelWeb

FunnelWeb is a literate programming tool, i.e., it supports production of **documents** that **also contain code**, and supports presentation of code in a sequence that is more appropriate for humans, while using macro-preprocessor-like functionality to arrange the "code chunks" in the right sequence for the target system of the code. The following example input file contains two *macro definitions* signalled by the *special sequence* "`@$`", and one *output file definition* (which can be seen as a special kind of macro definition) signalled by the special sequence "`@O`".

Inside the documentation, there can be references to code material, like the "`@{bin@}`" here in line 9; such references are delimited by the special sequences "`@{`" and "`@}`".

```
A frequently needed functionality is adding a
new first element to a search path environment variable:

@$@<addpath@>@(@2@)@M@{if [ -d @2 ]
then export @1=@2:$@1
fi@}

In many cases, the three standard paths for executables, man pages,
and (shared) libraries are extended with subdirectories @{bin@},
@{man@}, @{lib@} of a common installation prefix:

@$@<addpaths@>@(@1@)@M@{
if [ -d @1 ]
then
  @<addpath@>@(PATH@,@1/bin@)
  @<addpath@>@(MANPATH@,@1/man@)
  @<addpath@>@(LD_LIBRARY_PATH@,@1/lib@)
fi@}

We use this to create a file that will be ``sourced''
from our @{.bash_profile@}:

@O@<.bash_addpaths@>@{
@<addpaths@>@(/usr/local@)
@<addpaths@>@(/usr/local/packages/ghc-6.5@)
@}
```

For your information, the "literate documentation output" produced from this file by the original FunnelWeb is printed on page 2 (details of the graphical representation do not matter). An invocation of FunnelWeb on this example file will produce a file named .bash_addpaths; the contents of this is printed on 2, too.

In FunnelWeb, all *special sequences* start with the character "`@`" and have ***exactly one*** character after that; the special sequences for delimiting macro (and file) names are "`@<`" and "`@>`". For our simplified version, macro names consist of letters, digits, "`.`", "`_`", and spaces, but no newline characters.

Literate Documentation Output of FunnelWeb Example

A frequently needed functionality is adding a new first element to a search path environment variable:

addpath[1](◇2)**M** ≡
```
{if [ -d ◇2 ]
 then export ◇1=◇2:$◇1
 fi}
```
This macro is invoked in definitions 2, 2 and 2.

In many cases, the three standard paths for executables, man pages, and (shared) libraries are extended with subdirectories `bin`, `man`, `lib` of a common installation prefix:

addpaths[2](◇1)**M** ≡
```
{
 if [ -d ◇1 ]
 then
   addpath[1]('PATH','◇1/bin')
   addpath[1]('MANPATH','◇1/man')
   addpath[1]('LD_LIBRARY_PATH','◇1/lib')
 fi}
```
This macro is invoked in definitions 3 and 3.

We use this to create a file that will be "sourced" from our `.bash_profile`:

.bash_addpaths[3] ≡
```
{
 addpaths[2]('/usr/local')
 addpaths[2]('/usr/local/packages/ghc-6.3')
 }
```
This macro is attached to an output file.

FunnelWeb output file `.bash_addpaths`

```
if [ -d /usr/local ]
then
  if [ -d /usr/local/bin ]
  then export PATH=/usr/local/bin:$PATH
  fi
  if [ -d /usr/local/man ]
  then export MANPATH=/usr/local/man:$MANPATH
  fi
  if [ -d /usr/local/lib ]
  then export LD_LIBRARY_PATH=/usr/local/lib:$LD_LIBRARY_PATH
  fi
fi

if [ -d /usr/local/packages/ghc-6.5 ]
then
  if [ -d /usr/local/packages/ghc-6.5/bin ]
  then export PATH=/usr/local/packages/ghc-6.5/bin:$PATH
  fi
  if [ -d /usr/local/packages/ghc-6.5/man ]
  then export MANPATH=/usr/local/packages/ghc-6.5/man:$MANPATH
  fi
  if [ -d /usr/local/packages/ghc-6.5/lib ]
  then export LD_LIBRARY_PATH=/usr/local/packages/ghc-6.5/lib:$LD_LIBRARY_PATH
  fi
fi
```

(a) Give three regular expressions, one for each of the following three kinds of tokens:
– delimited macro names, as for example "`@<addpath@>`",
– special sequences occurring in the example, except "`@<`" and "`@>`", and
– text that does not contain any "`@`" characters.

Solution Hints

– `@<` [a-zA-Z1-9._]* `@>`

– `@` [$O()12M{},]

– [^@]+ **Note:** plus, not star — empty tokens are nonsense!

(b) Produce

- bison token type declarations and
- a flex lexer

for the token classes described in (a), except that **newline** characters should now be treated **as separate tokens**.

Solution Hints

```
%{
 #include <stdio.h>
 #include <stdbool.h>
 int yylex(void);
 void yyerror (char const * s) { fprintf(stderr, "%s\n", s); }
%}
%union {
 int     argnum;
 char *  string;

 // the remaining variants are for non-terminals:
 bool    multi;
 DefRec  macrodef;
 FileRec filedef;
};
%token <argnum> TOK_ARGREF
%token <string> TOK_TEXT
%token <string> TOK_MNAME
%token LPAREN RPAREN LBRACE RBRACE COMMA DOLLAR OUTPUT MULTI
```

```
%option noyywrap outfile="fw_lexer.c"
/* scanner for simplified FunnelWeb                    fw_lexer.l  */
%{
#include "fw_parser.tab.h"        /* token definitions and types */
%}
%%
```

```
"@<"[a–zA–Z0–1_\– ]+"@>"    yylval.string = strndup(yytext+2, strlen(yytext)–4); return TOK_MNAME;
"@"[1–9]                    yylval.argnum = atoi(yytext); return TOK_ARGREF;
"@("                        return LPAREN;
"@)"                        return RPAREN;
"@{"                        return LBRACE;
"@}"                        return RBRACE;
"@,"                        return COMMA;
"@$"                        return DOLLAR;
"@O"                        return OUTPUT;
"@M"                        return MULTI;
[^'@'\n]+                   yylval.string = strdup(yytext); return TOK_TEXT; // please ignore the quotes!
"\n"                        { return '\n';}
"@".                        fprintf(stderr, "Unrecognized special sequence: %s\n", yytext ); return –1;
.                           fprintf(stderr, "Unrecognized character: %s\n", yytext ); return –1;
```

The two macros in the example file above both have arguments, signalled by the "formal parameter header" "`@(@1@)`" for a macro with one argument and "`@(@2@)`" for a macro with two arguments; inside the macro body, "`@1`" stands for a reference to the first argument.

Code chunks can contain verbatim code, and also *macro invocations*, where macro names are again delimited by "`@<`" and "`@>`". Macros names in code chunks can be immediately followed by arguments in the shape of an *actual parameter list* delimited by "`@(`" and "`@)`"; if there are several arguments, these are separated by "`@,`" — all this is visible in the example. Each argument can again be an arbitrary code chunk.

Macro definition bodies are code chunks surrounded by the delimiters "`@{`" and "`@}`", just like code chunks embedded in documentation text.

The line structure of code chunks is relevant since all the lines in each product instance of a code chunk inherit the indentation level of its macro invocation — this effect can be seen in the product example on 2.

The opening "`@{`" of a macro definition body can be preceded by the special sequence "`@M`"; this indicates that this macro can be used several times.

(c) Give a **concrete** grammar in bison notation (for the time being without actions) for the aspects of FunnelWeb explained above such that your grammar reflects these explanations and the usage in the example.

Solution Hints

```
%type <macrodef> macrodef
%type <filedef>  filedef
%type <multi>  optmult
%type <argnum> optargheader

%start fw
%%

fw : /* empty */
  | fw docline '\n'
  | fw macrodef '\n'
  | fw filedef '\n'
```

```
docline : /* empty */
        | docline TOK_TEXT
        | docline body

body : LBRACE chunk RBRACE

filedef : OUTPUT TOK_MNAME body

macrodef : DOLLAR TOK_MNAME optargheader optmult body

optargheader : /* empty */              { $$ = 0; }
             | LPAREN TOK_ARGREF RPAREN  { $$ = $2; }

optmult : /* empty */  { $$ = false; }
        | MULTI        { $$ = true; }

chunkelem : TOK_TEXT
          | TOK_ARGREF
          | macrocall

chunkline : /* empty */
          | chunkline chunkelem

chunk : /* empty */
      | chunk chunkline '\n'

macroargs : chunk
          | macroargs COMMA chunk

macrocall : TOK_MNAME
          | TOK_MNAME  LPAREN  macroargs RPAREN
```

(d) Give an **abstract** grammar (for example in EBNF) for the aspects of FunnelWeb explained above.

Solution Hints

FW ::= (Doc | MacroDef | FileDef)*
Doc ::= (Text | Chunk)+
FileDef ::= Name Chunk
MacroDef ::= Name ArgNum Bool Chunk
CodeLine ::= (Text | ArgRef | MacroCall)*
Chunk ::= CodeLine*
MacroCall ::= Name Chunk*
ArgNum ::= [0-9]
ArgRef ::= [1-9]