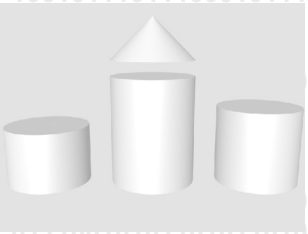

ODABA 8

General Purpose Classes

run Software-Werkstatt GmbH
Köpenicker Strasse 325
12555 Berlin

www.run-software.com

Tel: +49 (30) 65762791
Fax: +49 (30) 65762791
e-mail: run@run-software.com

Berlin, October 2003

Content

1 Introduction

General purpose classes (GPC) are service classes, which provide functionality for several services as handling lists, analyzing BNF based expressions or XML.

Platforms Service classes are defined as platform independent C++ classes for Windows platforms as well as for UNIX platforms (Linux, Solaris).

Open Source The GPC library is an open source library and can be used for commercial as well as for non commercial reasons.

Classes The document describes more complex GPCs and the way to use those. A detailed functional description is contained in the GPC function reference. This document describes the following features:

- BNFParser — Some classes are provided for defining expression by means of BNF, creating or generating BNF parsers based on a BNF definition and analyzing strings according to the defined syntax.
- XMLDocument — XML classes provide services for parsing or creating XML documents on a syntactical level.
- Client/Server — Client/Server classes are provided for client server communication based on TCP/IP. The classes provide a base for defining a more specific protocol based on TCP/IP. CS-classes support bi-directional communication as base for defining active servers.

2 BNF Parser

BNF classes are provided for defining expression by means of BNF, creating or generating BNF parsers based on a BNF definition and analyzing strings according to the defined syntax.

You may generate a C++ class for your parser or create an ad-hoc parser according to a given BNF. When parsing an expression the parser returns a syntax tree, that provides the values for the symbols found in the expression.

A BNF definition may refer to symbols defined in another BNF definition. This allows defining common BNF symbols e.g. for name and number (as in BNFStandardSymbols).

A string according to a given BNF syntax is based on a (top) BNF symbol. You may derive a specific BNF parsers for each type of BNF you want to support. The BNF is defined in the constructor for the BNF parser. Any number of spaces is allowed between symbols in a BNF but not required. Spaces are usually considered as separators between symbols.

Using BNF parser tools requires the following steps:

Defining a BNF Using parser classes requires a BNF definition that describes the syntax for the expressions to be parsed. Specific rules for defining a syntax are described in “Defining a BNF”.

Create parser From a given BNF definition you may create an ad-hoc parser or generate a C++ class for your parser definition. Creating an ad-hoc parser is good for testing the BNF, while generating a parser class can be considered as final step.

Analysing expressions Analysing expressions for a given syntax will create a BNF data tree, which contains nodes for each symbol found in the BNF.

Processing a BNFData tree The BNFData tree contains the nodes for the symbols found in the analyzed expression. You may list the BNFData tree or use several function for extracting the data for the nodes.

Defining a BNF

Using parser classes requires a BNF definition that describes the syntax for the expressions to be parsed. The rules for defining a BNF are described by a BNF, again, which is self-describing. The BNF described here includes some practical extensions as keywords and concurrency count.

BNF Syntax

The BNF syntax (i.e. the meta-BNF) is defined as follows:

```
  bnf          := bnf_stmt(*)
    bnf_stmt       :=  definition  |  keyword  |  reference  |
comment_line | nl

  definition   := sym_name ':=' rule [comment] nl
  rule         := prule [ alt_prule(*) ]
  alt_prule    := '|' prule
  prule        := ext_symbol(*)
  ext_symbol   := elm_symbol [ multiple ]
  multiple     := '(' maxnum ')'
  maxnum       := '*' | std_number
  elm_symbol   := sym_name | cstring | impl_symbol | opt_symbol
  opt_symbol   := '[' rule ']'
  impl_symbol  := '{' rule '}'
  sym_name     := name
  name         := std_name
  cstring      := std_string

  keyword      := sym_name '::' keydef [ alt_keydef(*) ] nl
  alt_keydef   := '|' keydef
  keydef       := cstring

  reference    := name '::=' symref nl
  symref       := class_ref | symbol_ref
  class_ref    := 'class' '(' name ')'
  symbol_ref   := 'ref' '(' name ')'

  comment_line:= comment nl
  comment      := '//' std_anychar(*)
  CC           := '//'

  std_symbols ::= class(BNFStandardSymbols)
  std_name    ::= ref(std_name)
  std_number  ::= ref(std_integer)
  std_string  ::= ref(std_string)
  std_anychar ::= ref(std_anychar)
  nl          ::= ref(std_nl)
```

You may define your own BNF specification, as long as you define the symbols with red bold letters. Other red symbols are optional and can be defined in your specific BNF definition (see “User-defined BNF Syntax”).

Separators

Blanks and tabs (9) are considered as separators between BNF symbols and must not be defined explicitly. New line characters (10,13) can be defined as automatic separators as well, but here, new line characters are used as symbols and not as separators, which allows terminating a BNF statement by new line characters.

Note, each BNF statement requires a line break at the end, i.e. the last statement must be followed by a line break as well, which results in an empty line of the definition file.

comments

A character sequence introducing a comment can be defined by the CC (comment characters) symbol. Any sequence beginning with this string at the beginning of a line or after any separated symbol (symbols that can be followed by one or more separators) is considered as comment until the line end. Thus, comments can be made at each end of line or as separate comment.

```
CC := '//'
```

The CC symbol is a reserved symbol and cannot be used otherwise.

bnf

A BNF definition contains a number of BNF statements, which are symbol definitions (symdef), symbol references (symref) or comments. Each statement is terminated by line break.

definition

A symbol definition defines one or more production rules (prule) for a symbol in the BNF or an empty line.

prule

A production rule for a symbol is a list of single or multiple symbols, which might be defined as optional.

alt_prule

Any number of alternative production rules may follow a production rule.

opt_symbol

Optional symbols are restricted in this specification to exactly one symbol. This, again, does not restrict the power of the BNF but requires for complex optional expressions the definition of a separate symbol.

mult_symbol

Multiple symbols is a small extension to standard BNF which allows defining a certain number or any number of symbols in a BNF definition. The following expression

```
x := symbol(*)
```

Defines any number of symbols and corresponds to

```
x := symbol [x]
```

Using multiple symbols has two advantages. One is, that it becomes much easier to define specific numbers of symbols as maximum 4. The other advantage is, that a multiple symbol appears as list in the BNF data tree, while the recursive definition would produce a hierarchy.

You may define multiple symbols as optional or not. Defining a multiple symbol like x(4) means, that x must appear exactly 4 times. Defining an optional multiple symbol like [x(4)] means, that x can appear maximum 4 times. Any number of symbols is indicated by '*' as number. x(*) means, that x must appear at least one time. [x(*)] means, that x may appear any number of times or not at all.

elm_symbol

Elementary symbols are basic elements of the rule. An elementary symbol may refer to a defined symbol, a string constant or an implicitly defined symbol.

sym_name

Symbol names must start with an alphabetical character and may contain numbers, '_' and '&' in the following characters

cstring

A string constant consists of one or more characters enclosed in '' ('+', '-', 'SELECT'). Quotes within a string constant can be defined with a preceding backslash (\'). String constants in a BNF expression acting as keywords. When being defined in a rule and not as keyword explicitly, the string constant is not reserved and can be used as e.g. name in other places.

Impl_symbol

Implicitly defined symbols are symbols, which do not get an explicit symbol name. Implicit symbols are defined as rule enclosed in { }.

keyword

Keywords are terminal symbols, which are reserved for specific use, only. Strings defined as keywords cannot be used in other roles within a document following the BNF rules.

```
_structure :: 'structure' | 'STRUCTURE'
```

The example above defines the ‘structure’ keyword. Then, ‘structure’ or ‘STRUCTURE’ cannot be used e.g. as name in the document (e.g. C++ file). One may, however, use ‘Structure’ as name, since keywords are case sensitive.

alt_keydef — Any number of string symbols can be defined for keywords.

keydef — A keyword definition defines the keyword string. Keyword strings must not contain spaces, tabs or line breaks.

reference — Symbol references can be used to refer to external symbol definitions in other BNF definitions.

class_ref — The BNF definition that contains the symbols to be referenced, must be referred to as class reference. Referring to an external BNF definition makes all symbols defined in the external definition available in the current definition.

symbol_ref — Specific symbols in the external BNF definition can be referenced by alias names to avoid naming conflicts for symbols.

Standard Symbols — The referenced BNF for standard symbols (BNFStandardSymbols) refers to the definition of common used BNF symbols. Standard symbols define specific character sets, numbers and name symbols as described in the subsequent BNF

```
std_constant  := std_float | std_string | std_bool

std_bool      := 'false' | 'true'

std_name      := std_alpha [ std_nchars ]

std_number    := std_integer | std_decimal | std_float
std_float     := std_decimal [ std_floatp ]
std_floatp    := 'E' std_integer
std_decimal   := std_integer [ std_decimalp ]
std_decimalp  := '.' std_digits
std_integer   := std_digits | '+' std_digits | '-' std_digits

std_string     := '\'' [std_str1(*)] '\'' | '"' [std_str2(*)]
'"'
std_str1      := std_cchar1(*) [ std_specc1 ] | std_specc1
std_specc1    := '\\' '\'' | std_bs
std_str2      := std_cchar2(*) [ std_specc2 ] | std_specc2
std_specc2    := '\\' '"' | std_bs

std_nchars    := std_nchar(*)
std_nchar     := std_alpha | std_digit | std_nspec
std_digits    := std_digit(*)
std_cchar1    := std_cchar | '"'
std_cchar2    := std_cchar | '\''
std_separators:= std_separator(*)
std_separator := ' ' | std_nl | 0x09
std_bs        := '\\' '\\'

std_nspec     := '_' | '$'
std_nl        := 0x0A | 0x0D 0x0A
std_digit     := 0 - 9[1]
std_alpha     := a - z | A - Z
std_cchar     := 1-255 except: ' " \
std_anychar   := 1-255 except: 0x0D, 0x0A
```

You may use alias names in you BNF to make it understandable, but you cannot use symbol names, which have already been defined in the standard BNF definition or in any other referenced BNF.

Ordering symbols

The order of symbols may play an important rule, when a symbol appears as starting symbol in several production rules. To avoid unlimited recursions and parser errors, some additional rules and suggestions for BNF definitions have been defined.

[1] This and the following BNF expressions conflict with the BNF syntax and are used here to make the definitions a little bit shorter

Completeness

All symbols referenced in the BNF definition (production rules) must be defined either in the BNF definition or in referenced external BNF definitions. Unresolved symbol references are shown when running the CreateParser function or when compiling the generated parser class.

Top-down

The BNF definition must be strict top-down, i.e. symbols should be defined after being referenced, or in other words: after defining a symbol it should not be referenced anymore.

It is not necessary to follow this rule in the BNF definition file, since the system will reorder the symbols according to this rule. When the BNF contains recursive definitions like:

```
a := b
b := a
```

which cannot be resolved, the CreateParser function or the BNF parser constructor will generate an error message and the BNF should not be used before solving the problem. You may create or generate a parser for recursive BNF definitions, but it may run into problems analyzing expressions defined by such a BNF.

Priority of symbols

The BNF parser assigns a priority to each symbol in the BNF definition according to the Top-down relationship between the symbols. When there are different possibilities for resolving an expression, symbols with higher priority are resolved before symbols with lower priority.

In some cases there are different ways for setting symbol priorities. In this case the implicit priority given in the BNF definition (first symbol highest priority, last symbol lowest) is used to determine the symbol priority.

Two symbols ahead

Since BNF definitions allow using symbols as starting symbols in several production rules, ambiguity cannot be avoided.

```
x := b
y := b c
```

To make the syntax as save as possible, more specific expressions should be defined before less specific ones (i.e. y should be defined before x in this case, because x is less specific since any symbol may follow b depending on the rest of the BNF definition). The parser is using a “looking two symbols ahead” mechanism, which guarantees, that expressions can be interpreted correctly as long as ambiguous production rules differ in the second symbol.

In this case, the latest symbol will get highest priority and will be evaluated before the previously defined symbol.

Top symbol

The first line in the BNF defines the top symbol, which gives the name to the BNF. When using the parser for analysing an expression it will always start with the top symbol, unless another symbol has been defined for analysing a sub-expression. The top symbol should not be referenced at any place in the BNF. When references become necessary, another symbol should be created:

```
expression := expr_def
expr_def   := a
a          := 'a' expr_def
```

(Note, that this definition is not recursive, since expr_def is not referenced as starting symbol.)

Example

In the following sections we will consider a simple BNF for arithmetical operations.

```
operation  := operand right_side(*)
right_side := operator operand
operand    := number | '(' operation ')'
operator   := '+' | '-' | '*' | '/'

std_symbols  ::= class(BNFStandardSymbols)
number       ::= ref(std_integer)
```

This is a simple BNF for defining any expression with the basic arithmetical operations. The expression refers to the standard definition for integer numbers as defined in the standard symbol BNF (BNFStandardSymbols).

User defined BNF syntax

There is not a real standard for writing a BNF. Nevertheless, most BNF notations are similar in principle, except the meta-symbols used in the BNF. Nevertheless, transforming a BNF into another “dialect” is a rather boring job. Hence, parser classes support alternative BNF definitions as long as the definition is based on the same symbols.

The following example shows a BNF notation used for defining the SQL-99 syntax:

```
bnf          := bnf_stmt(*)
bnf_stmt     := definition | comment_line | nl

definition   := sym_name '::=' rule nl nl
rule         := prule [ alt_prule(*) ]
alt_prule    := [nl] '|' [nl] prule
prule        := ext_symbol(*)
ext_symbol   := elm_symbol [ multiple ]
multiple     := '...'
elm_symbol   := sym_name | cstring | impl_symbol | opt_symbol
opt_symbol   := '[' rule ']'
impl_symbol  := '{' rule '}'
sym_name     := '<' name '>'
name         := std_name(*)
cstring      := std_string

cstring      := std_bnfchar(*) | string

CC           := '--'

std_symbols ::= class(BNFStandardSymbols)
std_name    ::= ref(std_name)
std_string  ::= ref(std_string)
std_bnfchar ::= ref(std_bnfchar)
nl          ::= ref(std_nl)
```

For running the syntax analysis with a user defined BNF syntax, you must define a path to the file containing the BNF definition.

In the example above, symbol names may consist of several names separated by blank. One may create a parser from such a BNF as well, but one cannot generate a C++ parser class from this definition, since symbol names are used in the parser class as variable names, which must not contain blanks.

Create Parser

After a BNF definition has been provided, this can be tested in two steps. The first step is creating the parser, which will report definition errors for the BNF definition file. From a given BNF definition you may create an ad-hoc parser or generate a C++ class for your parser definition. Creating an ad-hoc parser is good for testing the BNF, while generating a parser class can be considered as final step.

Build Parser

There are two ways of building a parser, which are rather likely.

```
#include <csos4mac.h>
#include <sBNFParser.hpp>
#include <sBNFData.hpp>

int main(int argc, char* argv[])
{
   BNFParser     *bparser;
   BNFData       *bdata = NULL;
   char          *path = "arop.bnf";
   char          *accpath = "arop.exp";

   GenerateParser(path,"e:/parser.cpp");
   if ( bparser = CreateParser(path,true) )
      bdata = bparser->AnalyzeFile(accpath,true);

// list BNF data tree structure
   if ( bdata )
     bdata->Print(0,YES);

   delete bdata; // delete bdata before deleting the parser!!!
   delete bparser;

  return(0);
}
```

The CreateParser function will print a symbol priority list on the console (list option in the CreateParser function):

```
Symbol list
    'arithmetical_operation'
    'operation'
    'right_side'
    'operand'
    'operator'
    'std_symbols'
    ... other std_sybols follow
```

Since operand and right_side have the same priority, the sequence in the BNF definition determines the priority ang gives operand a higher priority since it has been defined before right_side.

Generate parser class

Calling the GenerateParser() function as in the example above will generate a C++ parser class, which can be compiled immediately. This class may replace the ad-hoc server created from the BNF definition (CreateParser) in the example.

It is suggested creating and testing the parser before generating the parser class. The CreateParser() function allows passing a trace file name as third parameter, which will record all the attempts to resolve an expression. This is helpful for detecting errors in critical situations.

Check Expressions

In the second phase you can check several expressions with the parser created and view the result as BNFData tree. Calling the parser function Analyze as shown in the example above allows checking an expression. When successful, the function returns a BNF data tree, which can be displayed using the BNFData Print function.

Analyzing expressions

Analysing expressions for a given syntax will create a BNF data tree, which contains nodes for each symbol found in the BNF. You may analyse a complete syntax expression according to a given BNF definition but also a sub-expression.

For analysing a complete expression you may pass a filename or a 0-terminated string with the expression to the parser.

```
{
   BNFParser      *bparser = ...;
   BNFData        *bdata = NULL;

   ...

   bdata = bparser->AnalyzeFile(accpath,true);
   bdata = bparser->Analyze(string,true);

}
```

The skip option in the call (true) indicates, that the parser will skip separators at the beginning of the expression.

For analysing a sub-expression the parser must be called with the symbol name the sub-expression corresponds to. The sub-expression must be passed as 0-terminated string in this case.

```
{
   BNFParser      *bparser = ...;
   BNFData        *bdata = NULL;

   ...

   bdata = bparser->Analyze(string,"operand",true);

}
```

The symbol to be parsed is passed as symbol name. It must be a valid symbol defined for the parser or a referenced parser.

Since the parser always tries to analyse the complete expression, the string must not contain data after the end of the sub-expression. Otherwise the parser ill return an error.

Processing a BNFData tree

The result of analyzing an expression is a syntax tree, which consists of BNF data nodes. The syntax tree contains the nodes for the symbols found in the analyzed expression. You may list the BNFData nodes or use several function for extracting the data for the nodes.

Print syntax tree

The BNFData nodes in the syntax tree can be displayed using the BNFData Print function. The result for the expression

127 + 19 * (13 + 22) - (12/4)

analyzed according the the sample BNF arop.bnf will return the following syntax tree (each line represents a BNFData node):

```
arithmetical_operation: 127 + 19 * (13 + 22) - (12/4) ...
  arithmetical_operation: 127 + 19 * (13 + 22) - (12/4) ...
    operation: 127 + 19 * (13 + 22) - (12/4) ...
      operand: 127 ...
        std_integer: 127 ...
      right_side: + 19 ...
        operator: + ...
          operator: + ...
        operand: 19 ...
          std_integer: 19 ...
      right_side: * (13 + 22) ...
        operator: * ...
          operator: * ...
        operand: (13 + 22) ...
          operand: ( ...
          operation: 13 + 22 ...
            operand: 13 ...
              std_integer: 13 ...
            right_side: + 22 ...
              operator: + ...
                operator: + ...
   ... and so on
```

Each node in the tree is listed with the referenced symbol name and the value for the symbol.

Evaluate a syntax tree

There are several functions provided in the BNFData class that support browsing through the syntax tree. Iterator functions provide nodes on the same level, but you may also look for a specific symbol on a certain level or recursively. More details are available in the function reference (www.run-software.com/ODABADocu)

3 GenerateParser Utility

The GenerateParser utility allows creating a parser class for a specific BNF. This may reduce the analysing time since the parser need not to be built at runtime.

GenerateParser

The GenerateParser utility can be called with the following parameters:

GenerateParser *def_path*
[*cpp_path* [*trace_path* [*bnf_path*]]]

def_path

The definition path point to the location, where the BNF-file for the parser to be generated is stored.

cpp_path

The path refers to the location, where the generated C++-file will be stored. When the file does already exist, it will be replaced. When no cpp_path or NULL has been defined, the generated C++-file is stored at the same location as the def_file replacing the def_file extension with .cpp.

trace_path

When defining a trace path, the parsing steps of the process are recorded. This allows checking, what the parser has tried to analyse the def_file, which is important in some cases, when an error occurs. When no trace path or NULL is defined, no protocol is created.

bnf_path

A bnf_path can be passed to the function, when the BNF specification does not follow the standard definition of this BNF parser. This allows analysing imported BNF syntax, which may follow different rules. In this case, a BNF-definition as described in “User defined BNF syntax” must be provides at the location the bnf_path is pointing to.