Recursive Descent Parsing

Pat Morin

COMP 3002: Compiler Construction

Carleton
UNIVERSITY
Canada's Capital University

Tokenizing

- Tokenization is the process of converting input program text into a sequence of *tokens*
- A token usually has an ID and one or more attributes
- Examples:
  - <VARNAME, name="myVariable">
  - <FLOATCONST, value="1.3957">
  - <STRINGCONST, value="This is my string constant">
  - <WHILE>
  - <ASSIGN>, <EQ>, <LT>, <GT>, ...

Tokenizing Example

- The code snippet

```
while (x < 20)
  x := x + 2
```



```
<WHILE> <LPAREN> <VARNAME, "x"> <LT>
<INT, "20"> <RPAREN> <VARNAME, "x"> <ASSIGN>
<VARNAME, "x"> <PLUS> <INT, "2">
```

Carleton UNIVERSITY
Canada's Capital University

Purposes of Tokenizing

- Separates the program text into atomic *tokens*
  - This simplifies the parser code
  - Helps to separate the *grammar* from the *vocabulary*
- System-dependent details can be isolated
  - Reading from and buffering input
  - Handling different character encodings

Recursive Descent Parsing

- Recursive descent parsing is a method of writing a compiler as a collection of recursive functions
- This is usually done by converting a BNF grammar specification directly into recursive functions

Carleton
UNIVERSITY
Canada's Capital University

Backus-Naur Form

- *BNF* is a method of specifying the grammar of a programming language
- It is a list of rewriting rules
  - Each rule shows how the LHS can expand into a RHS

```
expression := <NUMBER> op expression
expression := <NUMBER>
expression := ( <expression> )
op := +
op := -
```

Carleton UNIVERSITY
Canada's Capital University

Recursive Descent Parser - Example

```
expression := <NUMBER>
expression := <NUMBER> op expression
op := +
op := -
```

```
ParseTreeNode expression() {
  Token n = getNumberToken();
  if ((Token o = getOpToken()) == null)
    return new NumberNode(n.value);
  }
  return new ExpressionNode(n.value,
                            o,
                            expression());
}
```

Carleton UNIVERSITY
Canada's Capital University

Recursive Descent Parser - Example

- Draw the parse tree (recursion tree) for this example:

$5 + 6 - 7 + 2$

Carleton UNIVERSITY
Canada's Capital University