jbailey

Bright lights shining upwards obscures the stars...

Using Antlr for Python

jbailey June 21st, 2008

I was looking for a Python lexer and parser, and found out that antlr can emit Python code with only minor twiddles of the input grammar.

The antlr site has a "Hello World" type of grammar for testing it out in Java: http://www.antlr.org/article/cutpaste/index.html (http://www.antlr.org/article/cutpaste/index.html)

What follows is that page ported to Python and Debian/Ubuntu. On a Debian-derived system:

```
$ sudo apt-get install antlr
```

Create a file called "t.g" with the following contents:

```
options {
        language=Python;
}

class P extends Parser;

startRule
    :   n:NAME
        {print "Hi there, "+n.getText()}
    ;

class L extends Lexer;

// one-or-more letters followed by a newline
NAME:   ( 'a'..'z'|'A'..'Z' )+ NEWLINE
    ;

NEWLINE
    :   '\r' '\n'   // DOS
    |   '\n'        // UNIX
    ;
```

Because antlr isn't in the CLASSPATH by default, the maintainer has included a "runantlr" script, which handles this for you. So,

```
$ runantlr t.g
```

This will create L.py, P.py and PTokenTypes.txt.

Now create main.py:

```
#!/usr/bin/python

import sys

import L
import P

l = L.Lexer(sys.stdin)
p = P.Parser(l)
p.startRule()
```

And run the resulting file:

```
$ python main.py
Jeff
Hi there, Jeff

$
```

(1 comment —)

: **sara_c:** That's pretty slick.

(1 comment —)