

## Homework 1

CS 440  
Spring 2004  
Craig A. Rich

Use the Generic Interpreter (gi) to build a lexical analyzer for the programming language Python, whose lexical specification is documented on the web at

<http://www.python.org/doc/current/ref/lexical.html>

A Python lexical analyzer should recognize the following tokens:

- SPACE
- COMMENT
- IDENTIFIER
- STRINGLITERAL
- INTEGER
- LONGINTEGER
- FLOATNUMBER
- IMAGNUMBER
- KEYWORD
- OPERATOR
- DELIMITER

The strings matching the tokens above are described in the Python lexical specification. The Python lexical specification also includes the tokens NEWLINE, INDENT and DEDENT, which we will not implement. We will consider them to be part of a more general white space token (SPACE) matched by strings of one or more white space characters. In Python, indentation is significant and is used to form implicit blocks (e.g., inside loops and conditional statements) without using any explicit delimiter symbols (e.g., curly braces). That's scary.

A sample Python source string is provided on the course web site

<http://www.csupomona.edu/~carich/classes/cs440/>

Please turn in a printed copy of your lexical analyzer source code, the output (including lexical debugging) produced by lexically analyzing the sample Python source provided, a Python source of your choosing containing at least one of each token, and the output produced when lexically analyzing it.