HOMEWORK#1:

My First Lexical Analyser

Due Date: **Friday, February 26th, 11:59.59pm**

Description:

Lexemes of a typical programming language are of different token types. Consider the following tokens:

- Integers are non-empty sequences of digits optionally preceded with either a '+' or '–' sign.
- Decimal numbers are Integers followed by a '.', followed by a non-empty sequence of digits. (e.g. `3.14`, `00.01`, `123.0`).
- Scientific numbers are Decimal numbers followed by the character '`E`', followed by a **non-zero** integer. (e.g. `12.0E4`, `1.23E-6`).
- Hexadecimal numbers are non-empty sequences of digits or the characters '`A`', '`B`', '`C`', '`D`', '`E`' or '`F`' followed by the suffix '`H`'. (e.g. `12AD0H`, `123H`, `1A2B3CH`,).
- Binary numbers are non-empty sequences of the digits '1' or '0' followed by the suffix '`B`'. (e.g. `10100B`, `101B`, `111110B`,).
- Keywords are specific strings that form the language. For this homework we will consider the the following keywords: '`while`', '`else`', '`end`', and '`print`'.
- Identifiers are strings that consists of a letter followed by zero or more letters, digits or the underscore; and that are **not** hexadecimal numbers (e.g. `x`, `size`, `name`, `p3`, `r_val`).

Write a Lexical Analyzer using the **automata encoding** techniques used in class to recognize Integers, Decimal numbers, Scientific numbers, Hexadecimal Numbers, Binary Numbers, Keywords and Identifiers.

Input:

The first line will be a positive integer **N**, followed by **N** strings to recognize, one per line.

Output:

The first line of the output should echo the number of input lines **N**. For every line of input, your program should output the line number and state if the string is recognized as either a Keyword, and Identifier, an Integer, a Decimal number, a Scientific number, Hexadecimal number, or an invalid string, in the format shown in the sample output.

Sample:

Input	Output
14 83462 -39874.454 while ABCH +234.34E-941 124.235.234 color -1.23E-3.5 4. +0 111B FFFF for4 3dfx	14 1: Integer. 2: Decimal. 3: Keyword. 4: Hexadecimal. 5: Scientific. 6: Invalid! 7: Identifier. 8: Invalid! 9: Invalid! 10: Integer. 11: Binary. 12: Identifier. 13: Identifier. 14: Invalid!

Submission:

Submit through the UNIX systems using the command '`cssubmit 3500 a 1`'.

If your submission is in C, C++, Java, Python 2 or Python 3, your main filename should be called '`mylexer.X`' where X is the extension appropriate to your programming language. { `.c`, `.cpp`, `.java`, `.py`, `.py3` }.

If you choose to submit in another language, make sure it runs in our Unix systems, and submit a `make` file that generates the necessary executable that will be tested using the command `./mylexer < testinput.txt`.

Hint:

```
#include <iostream>
#include <string>
using namespace std;

int main ()
{
    int T;
    string s;

    cin >> T;
    for (int i=0; i<T; i++){
        cin >> s;
        cout << "Hello " << s << "!" << endl;
    }
    return 0;
}
```