Last night I drifted off while reading a Lisp book.

At once, just like they said, I felt a great enlightenment. I saw the naked structure of Lisp code unfold before me.

Suddenly, I was bathed in a suffusion of blue.

The patterns and metapatterns danced. Syntax faded, and I swam in the purity of quantified conception. Of ideas manifest.

Truly, this was the language from which the gods wrought the universe.

No, it's not.

It's not?

I mean, ostensibly, yes. Honestly, we hacked most of it together with Perl.

Lisp is over half a century old and it still has this perfect, timeless air about it.

I wonder if the cycles will continue forever.

A few coders from each new generationrediscovering the Lisp arts.

These are your father's parentheses.

Elegant weapons for a more... civilized age.
CMSC 330: Organization of Programming Languages

Introduction

Instructor: Michael Lam (lam at cs)
TA: Richard Johnson (rbjohns8 at cs)
Course Website

http://www.cs.umd.edu/~lam/cmsc330/summer2012/

• Syllabus
• Office hours (TBD)
• Schedule
• Project assignments
• Resource links
• Web forum
• Submit server
Course Grading

- 5 projects (50%)
  - Late submission penalty (see syllabus)
- 5 quizzes (5%)
- Midterm + final exam (20/25%)

- Homeworks will not be graded and solutions will be posted online
Classroom reminders

• Arrive on time
• Sit up front (easier to see console text)
• Put cell phones on silenced or vibrate
• Laptops are encouraged, but be polite
• Interaction is good—ask questions!
• Attendance is neither mandatory nor directly rewarded
  – Lots of in-class examples
• Lectures will be available online
Project rules

• DO NOT SHARE PROJECT CODE!
• Re-using your own code is fine
• We will be using the submit server
  – Submit EARLY to ensure your setup works
• Most tests will be secret
  – Spend time writing your own test cases
  – Sharing your test cases is fine
• Project late penalties
  – 10% up to 24 hours late
  – 25% between 24 hours and 9am Friday
Other notes

• Take advantage of the web forum
• No textbook - use the lecture notes
• The summer semester is highly compressed
  – Stay organized and ahead of your work
• Get help as soon as you need it
  – Office hours will be posted online
• GLUE/GRACE accounts
  – glue.umd.edu/grace.umd.edu
  – UM directory ID+password
  – If you have a laptop, use it instead
Syllabus

- Scripting languages (Ruby)
  - Implicit vs. explicit declarations
  - Dynamic vs. static typing
  - Text processing and string manipulation

- Functional programming (Ocaml)
  - Lists and recursion
  - Higher-order programming and closures
  - Types and polymorphism
  - Data types and pattern matching
  - Modularity and abstraction
Syllabus

• Syntax
  • Regular expressions and finite automata
  • Context-free grammars

• Semantics
  • Lambda calculus
  • Evaluation strategies

• Implementation
  • Functions and parameter passing
  • Dynamic vs. static scoping
  • Concurrency
  • Generics and templates
  • Garbage collection and memory management
Course Goal

Learn how programming languages “work”

• Broaden your language horizons
  – Different programming languages
  – Different language features and tradeoffs
• Study how languages are implemented
  – What *really* happens when I write `x.foo(…)`?
• Study how languages are described
  – Mathematical formalisms
All Languages Are Equivalent

• A language is *Turing complete* if it can compute any function computable by a Turing Machine

• Essentially all general-purpose programming languages are Turing complete

• So why are there so many?
Why Study Programming Languages?

• Using the right language for a problem may be easier, faster, and less error-prone
  – Programming is a human activity
  – Features of a language make it easier or harder to program for a specific application
• To make you better at learning new languages
  – You may need to add code to a legacy system
    • E.g., FORTRAN (1954), COBOL (1959), …
  – You may need to write code in a new language
    • Your boss says, “From now on, all software will be written in {Ada/C++/Java/…}”
Why Study Programming Languages?

• To make you better at using languages you think you already know
  – Many “design patterns” in Java are functional programming techniques
  – Understanding what a language is good for will help you know when it is appropriate to use
Changing Language Goals

• 1950s-60s: Compile programs to execute efficiently
  – Language features based on hardware concepts
    • Integers, reals, goto statements
  – Programmers cheap; machines expensive
    • Keep the machine busy

• Today:
  – Language features based on design concepts
    • Encapsulation, records, inheritance, functionality, assertions
  – Processing power and memory very cheap; programmers expensive
    • Ease the programming process
Language Attributes to Consider

• Syntax -- What a program looks like

• Semantics -- What a program means

• Implementation -- How a program executes
Types of Languages

• Imperative/Procedural
• Declarative
  – Functional
  – Logical (rule-based)
• Object-oriented
• Scripting
• Data-oriented/Query
Imperative Languages

• Also called *procedural* or *von Neumann*
• Building blocks are functions and statements
  – Programs that write to memory are the norm
    
    ```c
    int x = 0;
    while (x < y) x := x + 1;
    ```

  – FORTRAN (1954)
  – Pascal (1970)
  – C (1971)
Functional Languages

• Also called \textit{applicative} languages
• No or few writes to memory
  – Functions are higher-order
    
    \begin{verbatim}
    let rec map f = function [] -> [] |
          x::l -> (f x)::(map f l)
    \end{verbatim}
  
  – LISP (1958)
  – ML (1973)
  – Scheme (1975)
  – Haskell (1987)
  – OCaml (1987)
Logical Languages

• Also called *rule-based* or *constraint-based*

• Program consists of a set of rules
  – “A :- B” – If B holds, then A holds
  • cat(tom).
  • animal(x) :- cat(x).

  – PROLOG (1970)
  – Various expert systems
Object-Oriented Languages

• Programs are built from objects
  – Objects combine functions and data
  – Often have classes and inheritance
  – “Base” may be either imperative or functional

```java
class C {
    int x;
    int getX() {return x;}
    ...
}
class D extends C {
    ...
}
```

– Smalltalk (1969)
– C++ (1986)
– OCaml (1987)
Scripting Languages

• Rapid prototyping languages for “little” tasks
  – Typically with rich text processing abilities
  – Generally very easy to use
  – “Base” may be imperative or functional; may be OO

```perl
#!/usr/bin/perl
for ($j = 0; $j < 2*$lc; $j++) {
    $a = int(rand($lc));
}
```

• sh (1971)
• perl (1987)
• Python (1991)
• Ruby (1993)
“Other” Languages

• There are lots of other languages around with various features
  – COBOL (1959) – Business applications
    • Imperative, rich file structure
  – BASIC (1964) – MS Visual Basic widely used
    • Originally an extremely simple language
    • Now a single word oxymoron
  – Logo (1968) – Introduction to programming
  – Forth (1969) – Mac Open Firmware
    • Extremely simple stack-based language for PDP-8
  – Ada (1979) – The DoD language
    • Realtime
  – Postscript (1982) – Printers- Based on Forth
    – …
Executing Programs

• Suppose we have a program $P$ written in a high-level language (i.e., not machine code)

• There are two main ways to run $P$
  - Compilation
  - Interpretation
Compilation (or “Translation”)

- Compiler translates code from one language to another
  - Result is a standalone executable
  - With optimization, can be highly efficient
**Interpretation**

- Interpreter executes each instruction one step at a time
  - Requires installed interpreter to run
  - Easier to port to new platforms
  - Usually less efficient (overhead of interpreter)
Steps of Compilation

for i < 6 do
  j += 10
next
Steps of Compilation: Lexing

1. Lexing/Scanning (Lexical Analysis) – Break up source code into tokens such as integer literals, identifiers, keywords, and operators
2. Parsing (Syntax Analysis) – Group tokens together into higher-level language constructs (conditionals, assignment statements, functions, …)
Steps of Compilation: Code Generation

3. Code Generation – Verify that the source program is valid and translate it into machine code or an internal “intermediate representation” (IR). There may be more than one intermediate representation for any given language.

```assembly
loop:
test %eax
jge done
mov *0x802c, %ecx
inc %ecx
mov %ecx, *0x802c
jmp loop
done:
```
Steps of Compilation: Optimization

4. Optimization (optional) – Improve the efficiency of the generated code
   – Eliminate dead code, redundant code, etc.
   – Change algorithm without changing functionality
     (e.g., $X=Y+Y+Y+Y \rightarrow X=4*Y \rightarrow X = Y$ shift left 2)
     • [If interested in compilation, take CMSC 430]
Compiler or Interpreter?

**gcc**
- Compiler – C code translated to object code, executed directly on hardware

**javac**
- Compiler – Java source code translated to Java byte code

**tcsh/bash**
- Interpreter – commands executed by shell program

**java**
- Interpreter – Java byte code executed by virtual machine
Compilation or Interpretation – Not so simple today

• Previously
  – Build program to use hardware efficiently
  – Often use of machine language for efficiency

• Today
  – No longer write directly in machine language
  – Use of layers of software
  – Concept of virtual machines
    • Each layer is a machine that provides functions for the next layer (e.g., javac/java distinction)
    • This is an example of *abstraction*, a basic building-block in computer science
Interactive Exercise

Introduce yourself to your neighbor(s) in groups of 2-4 and together write down as many languages as you can think of, specifying the following attributes:

Type: imperative, functional, or logical

Properties: object-oriented, scripting, or query

Execution: compiled, interpreted, or both
Attributes of a Good Language

• Clarity, simplicity, and unity
  • Provides both a framework for thinking about algorithms and a means of expressing those algorithms

• Orthogonality
  • Every combination of features is meaningful
  • Features work independently

What if, instead of working independently, adjusting the volume on your radio also changed the station? You would have to carefully change both simultaneously and it would become difficult to find the right station and keep it at the right volume. Your radio and tuning work orthogonally. And aren’t you glad they do!
Attributes of a Good Language

• Naturalness for the application
  • Program structure reflects the logical structure of algorithm

• Support for abstraction
  • Program data reflects problem being solved

• Ease of program verification
  • Verifying that program correctly performs its required function
Attributes of a Good Language

• Programming environment
  • External support for the language

• Portability of programs
  • Transportability of the resulting programs from the computer on which they are developed to other computer systems

• Cost of use
  • Program execution, program translation, program creation, and program maintenance
Who defines a language?

Is: \( I = 1 \land \land 2 + 3 \mid 4; \) legal in C?

- What is assigned to \( I \) if it is?
- Who makes this determination?

3 typical ways to answer this question:

1. Read a language book
   (Problem: Can you find one?)

2. Google it!
   (Problem: What exactly do you search for?)

3. Write a program to see what happens
   (Problem: Will it depend on whose compiler you use?)
Creation of standards

Language standards defined by national standards bodies:
- ISO - International Standards organization
- IEEE - Institute of Electrical and Electronics Engineers
- ANSI - American National Standards Institute

All work in a similar way:
1. Working group of volunteers set up to define standard
2. Agree on features for new standard
3. Vote on standard
4. If approved by working group, submitted to parent organization for approval
Creation of standards

• Standards in the US are voluntary:
  – There is no federal standards-making organization.
  – NIST - National Institute for Standards and Technology develops standards that are only required on federal agencies, not for commercial organizations.

• Consensus is the key to standards making:
  – Contentious features often omitted to gain consensus
  – Only vendors have a vested interest in the results
  – Users don't care until standard approved, and then it is too late!
Standards conforming programs

• Standards define behavior for a *standards conforming* program - one that meets the rules of the language standard

• In general (except for Ada), behavior of non-conforming program is not specified, so any extensions to a standards conforming compiler may still be standards conforming, even though the program is not standards conforming.

• Standards should be reviewed and updated periodically:
When to standardize a language?

- Problem: When to standardize a language?
  - If too late - many incompatible versions
    - FORTRAN in 1960s was already a de facto standard, but no two were the same
    - LISP in 1994, about 35 years after developed.
  - If too early - no experience with language - Ada in 1983 had no running compilers

- Other languages:
  - C in 1988
  - De facto standards: ML, SML, OCaml, Ruby
  - Smalltalk - none
  - Prolog - none
Internationalization

• Programming has become international
  – I18N issue - Internationalization - How to specify languages useful in a global economy?
• Character sets:
  – 1950s1960s – 6 bit sufficient (upper case, digits, special symbols …)
  – ASCII is a 7 bit 128 character code
  – Single 8-bit byte; usual format today - 256 character values. A lot in 1963, but insufficient today
• What about other languages?
  – Additional letters: German umlaut-ä, French accent-é, Scandanavian symbols-ö,
  – Russian, other alphabets (Greek, Arabic, Hebrew), ideographs (Chinese, Korean)?
  – Unicode - 16 bit code allows for 65K symbols. 8-bit byte is insufficient
Internationalization

• Some of the “I18N” issues:
  – What character codes to use?
  – Collating sequences? - How do you alphabetize various languages?
  – Dates? – If I said your exam was on 10/12/07 when would you show up?
• Time? - How do you handle time zones, summer time in Europe, or daylight savings time in US?
• Currency? - How to handle dollars, pounds, euros, etc.?
Summary

• Topics:
  – Syntax
  – Semantics
  – Implementation

• Types of languages:
  – Imperative
  – Functional
  – Logical
  – Object-oriented

• Execution methods:
  – Compilation
  – Interpretation

• Desirable language properties

• Standardization