CSE 373: Data Structure & Algorithms

Lecture 25: Programming Languages

Nicki Dell

Spring 2014

What is a Programming Language?

- A set of symbols and associated tools that translate (if necessary) collections of symbols into instructions to a machine
  - Compiler, execution platform (e.g. Java Virtual Machine)
  - Designed by someone or some people
    - Can have flaws, poor decisions, mistakes
- Syntax
  - What combinations of symbols are allowed
- Semantics
  - What those combinations mean
- These can be defined in different ways for different languages
- There are a lot of languages
  - Wikipedia lists 675 excluding dialects of BASIC and esoteric languages

Before High-Level Languages

- Everything done machine code or an assembly language
  - Arithmetic operations (add, multiply, etc.)
  - Memory operations (storing, loading)
  - Control operations (jump, branch)
- Example: move 8-bit value into a register
  - 1101 is binary code for move followed by 3-bit register id
  - 1101000 01100001
  - B0 61
  - MOV AL, 61h   ; Load AL with 97 decimal (61 hex)

Programming Language timeline

- C: 1973
- C++: 1980
- MATLAB: 1984
- Objective-C: 1986
- Mathematic (Wolfram): 1988
- Python: 1991
- Ruby: 1993
- Java: 1995
- Javascript: 1995
- PHP: 1995
- C#: 2001
- Scala: 2003

Choosing a Programming Language

- Most of the time you won't have a choice about what programming language to use
  - Software is already written in a particular language
  - Platform requires a specific language (Objective-C for iOS)
  - Language required by computational tool (Mathematica, etc.)
- Still important to understand capabilities and limitations of language
- When you do get to choose, your choice can have tremendous impact
  - This is despite theoretical equivalence!

Turing Completeness

- A programming language is said to be Turing complete if it can compute every computable function
  - Famous non-computable function is the Halting Problem
- So every Turing complete language can approximately simulate every other Turing complete language (do the same stuff)
- Virtually every programming language you might encounter is Turing complete
  - Data or markup languages (e.g. JSON, XML, HTML) are an exception
- So a choice of language is about how computation is described, not about what it's possible to compute

What we might want from a Language

- Readable (good syntax, intuitive semantics)
- High-level of abstraction (but still possible to access low level)
- Fast
- Good concurrency and parallelism
- Portable
- Manage side effects
- Expressive
- Make dumb things hard
- Secure
- Provably correct
- etc.

Type System

- Collection of rules to assign types to elements of the language
  - Values, variables, functions, etc.
- The goal is to reduce bugs
  - Logic errors, memory errors
  - Governed by type theory, an incredibly deep and complex topic

- The type safety of a language is the extent to which its type system prevents or discourages relevant type errors
  - Via type checking
- We'll cover the following questions:
  - When does the type system check?
  - What does the type system check?
  - What do we have to tell the type system?

When Does It Check? – Static type checking

- Static type-checking (check at compile-time)
  - Based on source code (program text)
  - If program passes, it's guaranteed to satisfy some type-safety properties on all possible inputs
  - Catches bugs early (program doesn't have to be run)
  - Possibly better run-time performance
    - Less (or no) checking to do while program runs
    - Compiler can optimize based on type
  - Not all useful features can be statically checked
    - Many languages use both static and dynamic checking

When Does it Check? – Dynamic type checking

- Dynamic type-checking (check at run-time)
  - Performed as the program is executing
  - Often “tag” objects with their type information
  - Look up type information when performing operations
  - Possibly faster development time
    - edit-compile-test-debug cycle
  - Fewer guarantees about program correctness

What Does it Check?

- Nominal type system (name-based type system)
  - Equivalence of types based on declared type names
  - Objects are only subtypes if explicitly declared so
  - Can be statically or dynamically checked
- Structural type system (property-based type system)
  - Equivalence of types based on structure/definition
  - An element A is compatible with an element B if for each feature in B's type, there's an identical feature in A's type
    - Not symmetric, subtyping handled similarly
- Duck typing
  - Type-checking only based on features actually used
  - Only generates run-time errors

How Much do we Have to Tell it?

- Type Inference
  - Automatically determining the type of an expression
  - Programmer can omit type annotations
    - Instead of (in C++)
      std::vector<int>::const_iterator itr = myvec.cbegin()
      use (in C++11)
      auto itr = myvec.cbegin()
  - Can make programming tasks easier
  - Only happens at compile-time
- Otherwise, types must be manifest (always written out)

How Flexible is it?

- Type conversion (typecasting)
  - Changing a value from one type to another, potentially changing the storage requirements
  - Reinterpreting the bit pattern of a value from one type to another
- Can happen explicitly or implicitly

```
double da = 3.3
double db = 3.3;
double dc = 3.4;
int result = (int)da + (int)db + (int)dc;
int result = da + db + dc;
```

What Does it All Mean?

- Most of these distinctions are not mutually exclusive
  - Languages that do static type-checking often have to do some dynamic type-checking as well
  - Some languages use a combination of nominal and duck typing
- Terminology is useful for describing language characteristics
- The terms “strong” or “weak” typing are often applied
  - These lack any formal definition
- Languages aren’t necessarily limited to “official” tools

Memory Safety

- Memory errors
  - Buffer overflow
  - Dynamic
  - Uninitialized variables
  - Out of memory
- Often closely tied to type safety
- Can be checked at compile-time or run-time (or not at all)
- Memory can be managed manually or automatically
  - Garbage collection is a type of automatic management
  - Some languages make use of both

Programming Paradigms

- A programming paradigm describes some fundamental way of constructing and organizing computer programs
  - A programming language supports one or more paradigms
- Imperative
  - A program is a series of statements that explicitly change the program state.
- Declarative
  - A program describes *what* should happen without describing *how* it happens
- Functional (can be considered a type of declarative)
  - Computation done by evaluation of functions, avoiding state and mutable data
- Object-oriented (as opposed to procedural)
  - Computation done via objects (containing data and methods)

Language Development

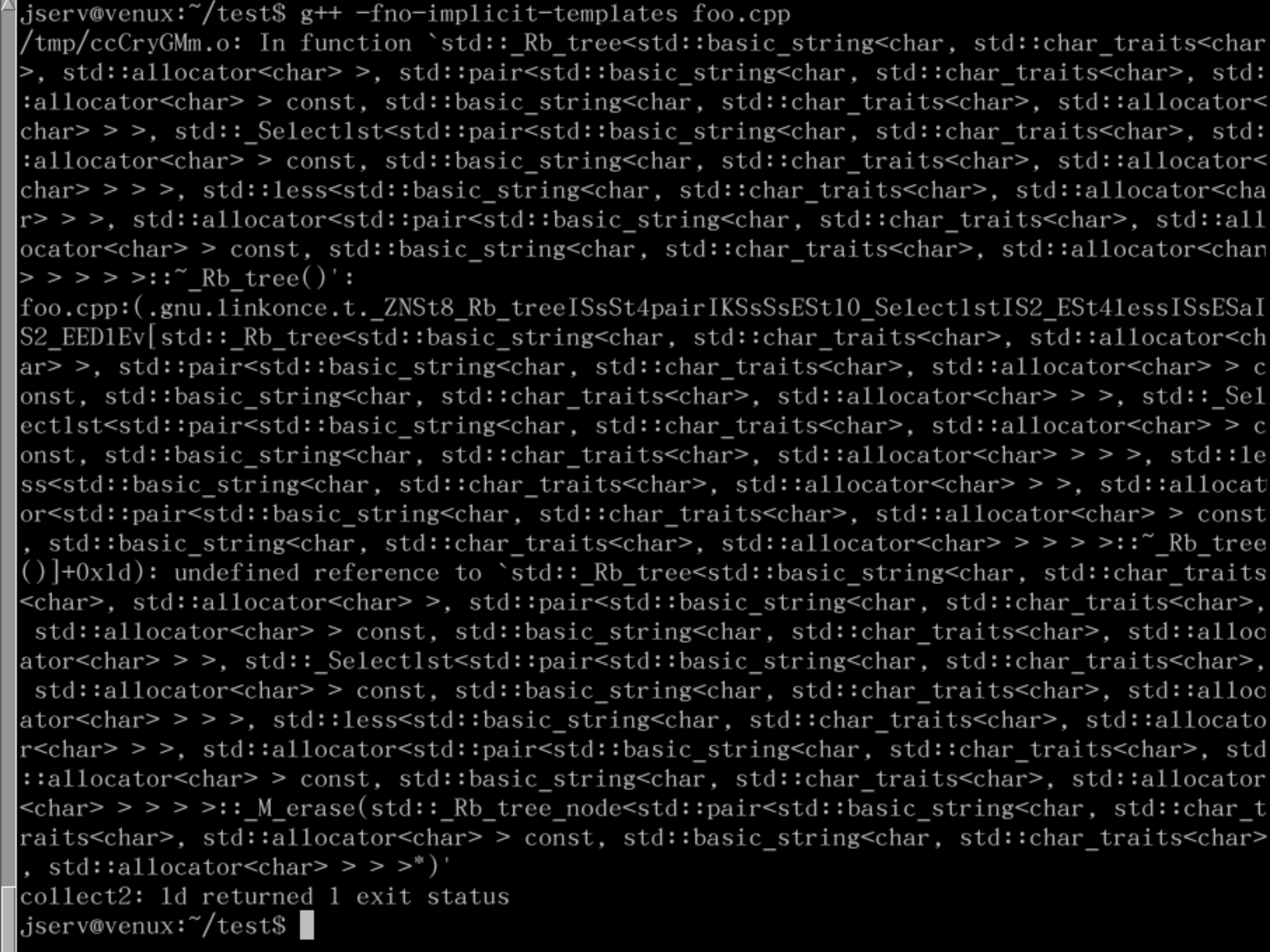
- Many attempts to develop a “universal language”
  - have failed due to diverse needs
  - program size, programmer expertise, program requirements, program evolution, and personal taste
- Languages often change over time
  - Generics were added to Java 9 years after initial release
  - Take extreme care not to break existing code
- One “standard,” many implementations
  - Standard defines syntax and semantics
- Whether a language will become popular is unpredictable
  - Some research suggests things like library availability and social factors may be more important than language features

Java

- Age: 19 years
- Developer: Oracle Corporation
- Paradigms: imperative, object-oriented
- Type system: static, nominative, manifest
- One of the most popular languages in use today
  - Lots of great tools and other resources
- Write Once, Run Anywhere approach (via JVM)
  - Used to be considered slow, improved by JIT optimization
  - Other languages using JVM (Scala, Jython, Clojure, Groovy)
- Can be quite verbose
- Sees lots of use in large-scale enterprise software
- I like Java (I've used it a lot). Some people hate java.

C/C++

- Age: 42/31 years
- Developer: International Organization for Standardization
- Paradigms: imperative, procedural, object-oriented (C++ only)
- Type system: static, nominative, manifest (C++11 has inference)
- Two of the most popular languages in use today
- "Closer to the hardware" than Java
  - Used where predictable resource use is necessary
  - OS, graphics, games, compilers
- Manual memory management, less protection from memory errors, sometimes inscrutable compiler errors
  - Generally easier to "do dumb things"
- I've used C/C++ for systems programming and for building graphics/vision applications.



```
jserv@venux:~/test$ g++ -fno-implicit-templates foo.cpp
/tmp/ccCryGMm.o: In function `std::_Rb_tree<std::basic_string<char, std::char_traits<char
>, std::allocator<char> >, std::pair<std::basic_string<char, std::char_traits<char>, std:
:allocator<char> > const, std::basic_string<char, std::char_traits<char>, std::allocator<
char> > >, std::_Select1st<std::pair<std::basic_string<char, std::char_traits<char>, std:
:allocator<char> > const, std::basic_string<char, std::char_traits<char>, std::allocator<
char> > > >, std::less<std::basic_string<char, std::char_traits<char>, std::allocator<cha
r> > >, std::allocator<std::pair<std::basic_string<char, std::char_traits<char>, std::all
ocator<char> > const, std::basic_string<char, std::char_traits<char>, std::allocator<char
> > > > >::~_Rb_tree()':
foo.cpp:(.gnu.linkonce.t._ZNSt8_Rb_treeISsSt4pairIKSsSsESt10_Select1stIS2_ESt4lessISsESaI
S2_EED1Ev[std::_Rb_tree<std::basic_string<char, std::char_traits<char>, std::allocator<ch
ar> >, std::pair<std::basic_string<char, std::char_traits<char>, std::allocator<char> > c
onst, std::basic_string<char, std::char_traits<char>, std::allocator<char> > >, std::_Sel
ect1st<std::pair<std::basic_string<char, std::char_traits<char>, std::allocator<char> > c
onst, std::basic_string<char, std::char_traits<char>, std::allocator<char> > > >, std::le
ss<std::basic_string<char, std::char_traits<char>, std::allocator<char> > >, std::allocat
or<std::pair<std::basic_string<char, std::char_traits<char>, std::allocator<char> > const
, std::basic_string<char, std::char_traits<char>, std::allocator<char> > > > >::~_Rb_tree
()]+0x1d): undefined reference to `std::_Rb_tree<std::basic_string<char, std::char_traits
<char>, std::allocator<char> >, std::pair<std::basic_string<char, std::char_traits<char>,
 std::allocator<char> > const, std::basic_string<char, std::char_traits<char>, std::alloc
ator<char> > >, std::_Select1st<std::pair<std::basic_string<char, std::char_traits<char>,
 std::allocator<char> > const, std::basic_string<char, std::char_traits<char>, std::alloc
ator<char> > > >, std::less<std::basic_string<char, std::char_traits<char>, std::allocato
r<char> > >, std::allocator<std::pair<std::basic_string<char, std::char_traits<char>, std
::allocator<char> > const, std::basic_string<char, std::char_traits<char>, std::allocator
<char> > > > >::_M_erase(std::_Rb_tree_node<std::pair<std::basic_string<char, std::char_t
raits<char>, std::allocator<char> > const, std::basic_string<char, std::char_traits<char>
, std::allocator<char> > > >*)'
collect2: ld returned 1 exit status
jserv@venux:~/test$
```

C#

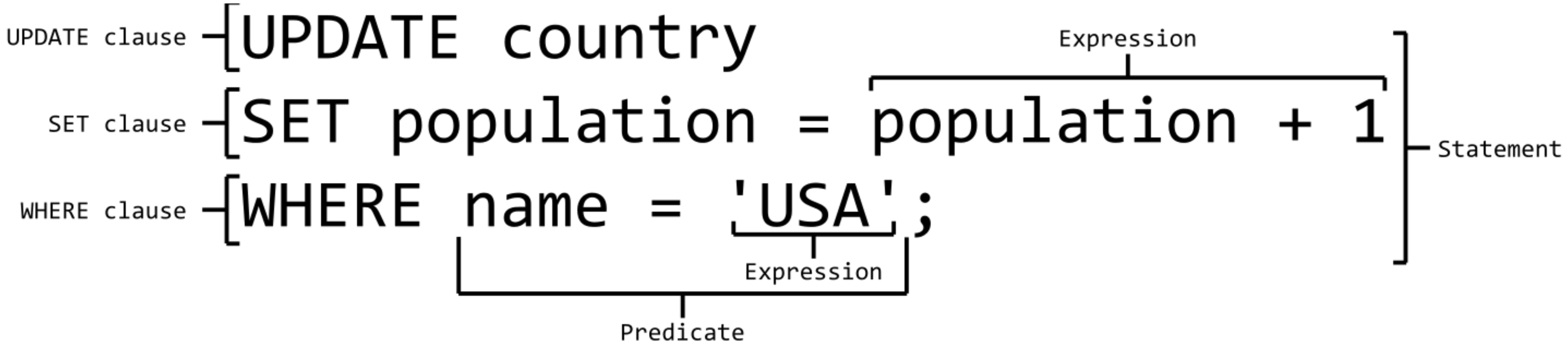
- Age: 14 years
- Developer: Microsoft
- Paradigms: imperative, object-oriented, functional
- Type system: static, nominative, partially inferred
  - optionally dynamic
- Runs on the .NET Framework
  - Provides things like garbage collection (similar to the JVM)
- Allows access to system functions with **`unsafe`** keyword
- Less verbose than Java, safer than C++
- Primary use is writing Windows applications
- I have programmed a little in C# (when at Microsoft), but Windows-only can be a big drawback

Haskell

- Age: 24 years
- Developer: many (research language)
- Paradigm: pure functional, lazy evaluation
- Type system: static, inferred
- Pure functional programming is a different way of thinking
  - maybe liberating, maybe frustrating
- Functional programming has seen only limited industrial use
- Safer and more transparent than an imperative language
  - Same function with same args always returns same value
  - Allows for compiler optimizations
- Performance suffers as hardware better suited to mutable data
- I think functional programming is fascinating, and enough languages include functional elements to make it worth learning

SQL (Structured Query Language)

- Age: 40 years
- Developer: ISO
- Paradigms: declarative
- Type system: static
- Used as a database query language
  - Declarative paradigm perfect for this application



```
UPDATE country
SET population = population + 1
WHERE name = 'USA';
```

- Using SQL is both easy and very powerful
- If you have a lot of data, definitely consider using free database software like MySQL

Python

- Age: 23 years
- Developer: Python Software Foundation
- Paradigm: imperative, object-oriented, functional, procedural
- Type system: dynamic, duck
- Has a Read-Eval-Print-Loop (REPL)
  - Useful for experimenting or one-off tasks
- Scripting language
  - Supports “scripts,” small programs run without compilation
- Often used in web development or scientific/numeric computing
- Variables don’t have types, only values have types
- Whitespace has semantic meaning
- Lack of variable types and compile-time checks mean more may be required of documentation and testing
- Python is my language of choice for accomplishing small tasks

JavaScript

- Age: 19 years
- Developer: Mozilla Foundation
- Paradigm: imperative, object-oriented, functional, procedural
- Type system: dynamic, duck
- Also a scripting language (online/browser REPLs exist)
- Primary client-side language of the web
- Takes a continue at any cost approach
  - Shared by many web-focused languages (PHP, HTML)
  - Things that would be errors in other languages don't stop execution, and are allowed to fail silently
- JavaScript is nice for simple things, immediately running on the web is great, but doing larger/more complex software is terrible

PHP

- Age: 19 years
- Developer: The PHP Group
- Paradigm: imperative, object-oriented, functional, procedural
- Type system: dynamic
- Works with Apache (>50% all websites), so very common server-side language
- Minimal type system, lots of strange behavior, just awful
  - If two strings are compared with ==, PHP will silently cast them to numbers (0e45h7 == 0w2318 evaluates to true)
- I hope none of you will ever have to use (or choose to use) PHP