Data Types

- most programming languages include a notion of types for expressions, variables, and values
- two main purposes of types
  - imply a context for operations
    - programmer does not have to state explicitly
    - example: $a + b$ means mathematical addition if $a$ and $b$ are integers
  - limit the set of operations that can be performed
    - prevents the programmer from making mistakes
    - type checking cannot prevent all meaningless operations
    - it catches enough of them to be useful
Type Systems

• bits in memory can be interpreted in different ways
  – instructions
  – addresses
  – characters
  – integers, floats, etc.

• bits themselves are untyped, but high-level languages associate types with values
  – useful for operator context
  – allows error checking
Type Systems

• a type system consists of
  – a way to define types and associate them with language constructs
    • constructs that must have values include constants, variables, record fields, parameters, literal constants, subroutines, and complex expressions containing these
  – rules for type equivalence, type compatibility, and type inference
    • type equivalence: when the types of two values are the same
    • type compatibility: when a value of a given type can be used in a particular context
    • type inference: type of an expression based on the types of its parts
Type Checking

• type checking ensures a program obeys the language’s type compatibility rules
  – type clash: violation of type rules

• a language is strongly typed if it prohibits an operation from performing on an object it does not support

• a language is statically typed if it is strongly typed and type checking can be performed at compile time
  – few languages fall completely in this category
  – questions about its value
Type Checking

• examples
  – Common Lisp is strongly typed, but not statically typed
  – Ada is statically typed
  – Pascal is almost statically typed
  – Java is strongly typed, with a non-trivial mix of things that can be checked statically and things that have to be checked dynamically
Type Checking

• a language is dynamically typed if type checking is performed at run time
  – late binding
  – List and Smalltalk
  – scripting languages, such as Python and Ruby
Polymorphism

- Polymorphism results when the compiler finds that it doesn't need to know certain things
  - allows a single body of code to work with objects of multiple types
  - with dynamic typing, arbitrary operations can be applied to arbitrary objects
    - implicit parametric polymorphism: types can be thought of to be implied unspecified parameters
    - incurs significant run-time costs
    - for example, ML’s type inference/unification scheme
Polymorphism

- subtype polymorphism in object-oriented languages
  - a variable of the base type can refer to an object of the derived type
  - explicit parametric polymorphism, or generics
    - C++, Eiffel, Java
    - useful for container classes
      List<T> where T is left unspecified
we all have developed an intuitive notion of what types are; what's behind the intuition?

- collection of values from a domain (the denotational approach)
- internal structure of data, described down to the level of a small set of fundamental types (the structural approach)
- collection of well-defined operations that can be applied to objects of that type (the abstraction approach)
Data Types

• specified in different ways
  – Fortran: spelling of variable names
  – Lisp: infer types at compile time
  – List, Smalltalk: track at run time
  – C, C++: declared at compile time
Classification of Types

- most languages provide a set of built-in types
  - integer
  - boolean
    - often single character with 1 as true and 0 as false
    - C: no explicit boolean; 0 is false, anything else is true
  - char
    - traditionally one byte
    - ASCII, Unicode
  - real
Classification of Types

- numeric types
  - C and Fortran distinguish between different lengths of integers
  - C, C++, C#, Modula-2: signed and unsigned integers
  - differences in real precision cause unwanted behavior across machine platforms
  - some languages provide complex, rational, or decimal types
Classification of Types

- enumeration types
  - facilitate readability
  - allow compiler to catch errors

- Pascal example

```pascal
type weekday = (sun, mon, tue, wed, thu, fri, sat);

for today := mon to fri do begin ...

var daily_attendance : array [weekday] of integer;
```
Classification of Types

• alternatively, can be declared as constants

\[
\text{const sun} = 0; \text{mon} = 1; \text{tue} = 2; \text{wed} = 3; \text{thu} = 4; \text{fri} = 5; \text{sat} = 6;
\]

– same as C

\[
\text{enum weekday \{sun, mon, tue, wed, thu, fri, sat\};}
\]
Classification of Types

- subrange types
  - contiguous subset of discrete base type

```plaintext
type test_score = 0..100;
workday = mon..fri;
```

- helps to document code
- most store the actual values rather than ordinal locations

```plaintext
type water_temperature = 273..373; (* degrees Kelvin *)
```

- needs two bytes
Classification of Types

• composite types
  – records
    • Cobol
    • fields of possibly different type
    • similar to tuples
  – variant records (unions)
    • multiple field types, but only one valid at a given time
  – arrays
    • aggregate of same type
    • strings
Classification of Types

• composite types (cont.)
  – sets
    • discrete
    • like enumerations and subranges
  – pointers
    • good for recursive data types
  – lists
    • head element and following elements
    • variable length, no indexing
  – files
    • mass storage, outside program memory
    • linked to physical devices (e.g., sequential access)
Orthogonality

- orthogonality is a useful goal in the design of a language, particularly its type system
  - a collection of features is orthogonal if there are no restrictions on the ways in which the features can be combined (analogy to vectors)
Orthogonality

• for example
  – Pascal is more orthogonal than Fortran, (because it allows arrays of anything, for instance), but it does not permit variant records as arbitrary fields of other records (for instance)

• orthogonality is nice primarily because it makes a language easy to understand, easy to use, and easy to reason about
Type Checking

- A type system has rules for:
  - type equivalence (when are the types of two values the same?)
  - type compatibility (when can a value of type A be used in a context that expects type B?)
  - type inference (what is the type of an expression, given the types of the operands?)
Type Checking

• type compatibility / type equivalence
  – compatibility is the more useful concept, because it tells you what you can do
  – terms are often used interchangeably (incorrectly)
Type Checking

- format does not matter
  
  \[
  \text{struct \{ int a, b; \}}
  \]

  \text{is the same as}
  
  \[
  \text{struct \{}
  
  \quad \text{int a, b;}
  
  \}
  \]

- want them to be the same as
  
  \[
  \text{struct \{}
  
  \quad \text{int a;}
  
  \quad \text{int b;}
  
  \}
  \]
Type Checking

- two major approaches: structural equivalence and name equivalence
  - name equivalence is based on declarations
  - structural equivalence is based on some notion of meaning behind those declarations
  - name equivalence is more fashionable these days
Type Checking

• at least two common variants on name equivalence
  – differences between all these approaches boils down to where the line is drawn between important and unimportant differences between type descriptions
  – in all three schemes, we begin by putting every type description in a standard form that takes care of "obviously unimportant" distinctions like those above
Type Checking

• structural equivalence depends on simple comparison of type descriptions
  – substitute out all names
  – expand all the way to built-in types
• original types are equivalent if the expanded type descriptions are the same
Type Checking

• example in Ada illustrating type conversion

n : integer;  -- assume 32 bits
r : real;  -- assume IEEE double-precision
t : test_score;  -- as in Example 7.9 (1..100)
c : celsius_temp;  -- as in Example 7.19 (integer)

...  
t := test_score(n);  -- run-time semantic check required
n := integer(t);  -- no check req.; every test_score is an int
r := real(n);  -- requires run-time conversion
n := integer(r);  -- requires run-time conversion and check
n := integer(c);  -- no run-time code required
c := celsius_temp(n);  -- no run-time code required
Type Checking

• coercion
  – when an expression of one type is used in a context where a different type is expected, one normally gets a type error
  – but what about
    
    ```
    var a : integer; b, c : real;
    ...
    c := a + b;
    ```
Type Checking

• coercion
  – many languages allow such statements, and coerce an expression to be of the proper type
  – coercion can be based just on types of operands, or can take into account expected type from surrounding context as well
  – Fortran has lots of coercion, all based on operand type
Type Checking

• C has lots of coercion, too, but with simpler rules:
  – all floats in expressions become doubles
  – short int and char become int in expressions
  – if necessary, precision is removed when assigning into LHS
Type Checking

• example in C, which does a bit of coercion

```c
short int s;
unsigned long int l;
char c;     /* may be signed or unsigned -- implementation-dependent */
float f;    /* usually IEEE single-precision */
double d;   /* usually IEEE double-precision */
...
s = l;      /* l’s low-order bits are interpreted as a signed number. */
l = s;      /* s is sign-extended to the longer length, then
    its bits are interpreted as an unsigned number. */
s = c;      /* c is either sign-extended or zero-extended to s’s length;
    the result is then interpreted as a signed number. */
f = l;      /* l is converted to floating-point. Since f has fewer
    significant bits, some precision may be lost. */

f = d;      /* f is converted to the longer format; no precision lost. */
f = d;      /* d is converted to the shorter format; precision may be lost.
    If d’s value cannot be represented in single-precision, the
    result is undefined, but NOT a dynamic semantic error. */
```
Type Checking

• in effect, coercion rules are a relaxation of type checking
  – recent thought is that this is probably a bad idea
  – languages such as Modula-2 and Ada do not permit coercions
  – C++, however, goes hog-wild with them
  – they're one of the hardest parts of the language to understand
Type Checking

• make sure you understand the difference between
  – type conversions (explicit)
  – type coercions (implicit)
  – sometimes the word 'cast' is used for conversions (e.g., C)
Type Checking

- universal reference types
  - a reference that can point to anything
    - C, C+: void
    - Clu: any
    - Modula2: address
    - Java: Object
    - C#: object
• records
  – allow related heterogeneous types to be stored together
  – C, C++, and Common Lisp: structure
    • C++: class with globally visible members
  – Fortran: types
Records (Structures) and Variants (Unions)

• C

```c
struct element {
    char name[2];
    int atomic_number;
    double atomic_weight;
    _Bool metallic;
};
```

• Pascal

```pascal
type two_chars = packed array [1..2] of char;
(* Packed arrays will be explained in Example 7.39. Packed arrays of char are compatible with quoted strings. *)

type element = record
    name : two_chars;
    atomic_number : integer;
    atomic_weight : real;
    metallic : Boolean
end;
```
• each record component is known as a field
  – most languages use dot notation to refer to fields

```plaintext
var copper : element;
const AN = 6.022e23;  (* Avogadro’s number *)
...
copper.name := 'Cu';
atoms := mass / copper.atomic_weight * AN;
```

– Fortran: copper%name
Records (Structures) and Variants (Unions)

- most languages allow records to be nested

```haskell
type short_string = packed array [1..30] of char;
type ore = record
  name : short_string;
  element_yielded : record
    name : two_chars;
    atomic_number : integer;
    atomic_weight : real;
    metallic : Boolean
  end
end;
```

- alternatively

```haskell
type ore = record
  name : short_string;
  element_yielded : element
end;
```
Records (Structures) and Variants (Unions)

- both reference
  - `malachite.element_yielded.atomic_number`

- in C, struct tags appear in struct definitions as

```c
struct element {
    char name[2];
    int atomic_number;
    double atomic_weight;
    _Bool metallic;
};
```
  - in C, need typedef
    - typedef struct ...
  - in C++, struct tag is type name
Records (Structures) and Variants (Unions)

- records
  - usually laid out contiguously
    - addresses are computed as base + offset
  - possible holes for alignment reasons
  - smart compilers may re-arrange fields to minimize holes (C compilers promise not to)
  - implementation problems are caused by records containing dynamic arrays
Records (Structures) and Variants (Unions)

- records
  - example
  - holes for alignment

<table>
<thead>
<tr>
<th>name</th>
<th>atomic_number</th>
<th>atomic_weight</th>
<th>metallic</th>
</tr>
</thead>
</table>

4 bytes/32 bits
Records (Structures) and Variants (Unions)

- **records**
  - some languages, such as Pascal, allow the data to be packed
    - compiler should optimize for space over speed

```pascal
type element = packed record
  name : two_chars;
  atomic_number : integer;
  atomic_weight : real;
  metallic : Boolean
end;
```
Records (Structures) and Variants (Unions)

- records
  - packed layout

<table>
<thead>
<tr>
<th>name</th>
<th>atomic_</th>
</tr>
</thead>
<tbody>
<tr>
<td>number</td>
<td></td>
</tr>
<tr>
<td>atomic_weight</td>
<td></td>
</tr>
<tr>
<td>metallic</td>
<td></td>
</tr>
</tbody>
</table>

- less space, but system takes longer to access elements
• records
  – some languages, such as Ada, allow record assignment
    • my_element := copper
  – or record tests for equality
    • if my_element = copper then
  – most other languages do not
Records (Structures) and Variants (Unions)

- records
  - record equality can be done field by field
  - to save time, a block compare function can be used
    - compares bytes
    - problems?
      - holes need to be filled with 0’s to allow block comparisons to work correctly
• records
  – holes in records waste space
    • packing helps, but with performance penalties
    • smart compilers rearrange fields automatically (not C compilers)
    • programmer unaware
    • some systems-level applications may depend on non-rearranged fields

![Diagram showing records and their fields]

Copyright © 2005 Elsevier
Records (Structures) and Variants (Unions)

• unions (variant records)
  – space was extremely limited in early days of programming
  – overlay space
  – cause problems for type checking

• can be used for automatically re-interpreting bits without casts
Records (Structures) and Variants (Unions)

- Pascal

```pascal
type long_string = packed array [1..200] of char;
type string_ptr = ^long_string;
type element = record
  name : two_chars;
  atomic_number : integer;
  atomic_weight : real;
  metallic : Boolean;
  case naturally_occuring : Boolean of
    true : ( 
      source : string_ptr;
      (* textual description of principal commercial source *)
      prevalence : real;
      (* fraction, by weight, of Earth’s crust *)
    );
    false : ( 
      lifetime : real;
      (* half-life in seconds of the most stable known isotope *)
    )
  end;
```
Records (Structures) and Variants (Unions)

- variant part must be at end
• Modula-2: variant part not required to be at end
Arrays

- arrays are the most common and important composite data types
- unlike records, which group related fields of disparate types, arrays are usually homogeneous
- semantically, they can be thought of as a mapping from an *index type* to a *component* or *element type*
Arrays

• index
  – most languages require the index to be an integer
  – or at least discrete type
  – non-discrete type
    • associative array
    • Python: dictionary
    • require hash table for lookup
Arrays

- syntax and operation
  - index usually inside square brackets [ ]
  - some languages use parentheses ( )
Arrays

• declaration
  – C
    ```c
    char upper[26];
    ```
  – Fortran
    ```fortran
    character, dimension (1:26) :: upper
    character (26) upper       ! shorthand notation
    ```
  – Pascal
    ```pascal
    var upper : array ['a'..'z'] of char;
    ```
  – Ada
    ```ada
    upper : array (character range 'a'..'z') of character;
    ```
Arrays

• multidimensional arrays

  matrix : array (1..10, 1..10) of real;  -- Ada
  real, dimension (10,10) :: matrix       ! Fortran

  – can be accessed as matrix [3][4]
  – sometimes as matrix [3, 4]
  – Ada: matrix (3, 4)
Arrays

- a slice or section is a rectangular portion of an array

- matrix(3:6, 4:7)
- matrix(6:, 5)
- matrix(:, 2:8:2)
- matrix(:, [2, 5, 9])
Arrays

• slicing
  – some languages allow operations on arrays
    • Ada: lexicographic ordering: if A < B
    • Fortran: more than 60 intrinsic functions
      – arithmetic operations, logic functions, bit manipulation, trigonometric functions
Arrays

- dimensions, bounds, and allocation
  - global lifetime, static shape — if the shape of an array is known at compile time, and if the array can exist throughout the execution of the program, then the compiler can allocate space for the array in static global memory
  - local lifetime, static shape — if the shape of the array is known at compile time, but the array should not exist throughout the execution of the program, then space can be allocated in the subroutine’s stack frame at run time.
  - local lifetime, shape bound at elaboration time
Arrays

• dope vector
  – collection of information about an array
    • dimensions
    • bounds (lower and upper)
  – record
    • offset to each field
  – good for runtime checks
  – can enhance efficiency by avoiding computations
  – origin: “having the dope on (something)”
Arrays

- stack allocation
  - conformant arrays
    - parameters to subroutines
    - dimensions not known
    - dope vector passed to help

```pascal
function DotProduct(A, B : array [lower..upper : integer] of real) : real;
var i : integer;
  rtn : real;
begin
  rtn := 0;
  for i := lower to upper do rtn := rtn + A[i] * B[i];
  DotProduct := rtn
end;
```
Arrays

procedure foo (size : integer) is
M : array (1..size, 1..size) of real;
...
begin
...
end foo;

Figure 7.7: Allocation in Ada of local arrays whose shape is bound at elaboration time. Here \( M \) is a square two-dimensional array whose width is determined by a parameter passed to \texttt{foo} at run time. The compiler arranges for a pointer to \( M \) to reside at a static offset from the frame pointer. \( M \) cannot be placed among the other local variables because it would prevent those higher in the frame from having static offsets. Additional variable-size arrays are easily accommodated. The purpose of the \textit{dope vector} field is explained in Section 7.4.3.
Arrays

- heap allocation
  - dynamic arrays
  - dimensions not known until runtime
  - dope vector can be maintained dynamically
Arrays

• arrays are stored in contiguous memory locations

• for multidimensional arrays, an ordering of dimensions must be specified
  – column-major - only in Fortran
  – row-major
    • used by everybody else
  – important for element access
Arrays

Row-major order

Column-major order
Arrays

• to access elements in a single row
  – column-major
    • cache may not be big enough to store next row item
  – row-major
    • effective
  – reversed for column support, but row support seems to be used more often
Arrays

- two layout strategies for arrays
  - contiguous elements
  - row pointers
- row pointers
  - an option in C
  - allows rows to be put anywhere - nice for big arrays on machines with segmentation problems
  - avoids multiplication
  - nice for matrices whose rows are of different lengths
    - e.g. an array of strings
  - requires extra space for the pointers
Arrays

```c
char days[][10] = {
    "Sunday", "Monday", "Tuesday",
    "Wednesday", "Thursday",
    "Friday", "Saturday"
};
...
days[2][3] == 's'; /* in Tuesday */
```

```c
char *days[] = {
    "Sunday", "Monday", "Tuesday",
    "Wednesday", "Thursday",
    "Friday", "Saturday"
};
...
days[2][3] == 's'; /* in Tuesday */
```
Arrays

- address calculation
  - can compute some portions before execution
  - declaration

\[
A : \text{array } [L_1 \ldots U_1] \text{ of array } [L_2 \ldots U_2] \text{ of array } [L_3 \ldots U_3] \text{ of elem\_type};
\]

- constants

\[
S_3 = \text{size of elem\_type}
\]
\[
S_2 = (U_3 - L_3 + 1) \times S_3
\]
\[
S_1 = (U_2 - L_2 + 1) \times S_2
\]

- address

address of \( A \)
\[
+ (i - L_1) \times S_1
\]
\[
+ (j - L_2) \times S_2
\]
\[
+ (k - L_3) \times S_3
\]
Arrays
Arrays

• we can compute the address with fewer operations

\[(i \times S_1) + (j \times S_2) - (L_1 \times S_1) + \text{address of } A - [(L_2 \times S_2) + (L_3 \times S_3) - (3 \times S_3)]\]

  – the calculations in square brackets is compile-time constant that depends only on the type of A
Strings

- strings are really just arrays of characters
- they are often special-cased, to give them flexibility (like polymorphism or dynamic sizing) that is not available for arrays in general
  - easier to provide these operations for strings than for arrays in general because strings are one-dimensional and (more importantly) non-circular
Sets

• possible implementations
  – bitsets are what usually get built into programming languages
  – intersection, union, membership, etc. can be implemented efficiently with bitwise logical instructions
  – some languages place limits on the sizes of sets to make it easier for the implementer
    • for 32-bit int’s, bitset size is 500MB
    • for 64-bit int’s, bitset size enormous
Points and Recursive Types

- Pointers serve two purposes
  - Efficient (and sometimes intuitive) access to elaborated objects (as in C)
  - Dynamic creation of linked data structures (recursive types), in conjunction with a heap storage manager

- Several languages (e.g., Pascal) restrict pointers to accessing objects in the heap
Pointers and Recursive Types

• operations on pointers
  – allocation and deallocation of objects in the heap
  – dereferencing of pointers to access the objects to which they point
  – assignment from one pointer to another

• operations behave differently depending on whether the language employs a reference model or value model for names
  – functional languages use reference model
Pointers and Recursive Types

- reference model
  - ML example
    
    ```
    datatype chr_tree = empty | node of char * chr_tree * chr_tree;
    ```
  - example tree
    
    ```
    (#"R", node (#"X", empty, empty), node (#"Y", node (#"Z", empty, empty), node (#"W", empty, empty)))
    ```
  - block storage from heap
Pointers and Recursive Types

Diagram:

- Node R
- Node X
- Node Y
- Node Z
- Node W
- Empty
Pointers and Recursive Types

• reference model
  – Lisp example
    ’(#\R (#\X ()()) (#\Y (#\Z ()())(#\W ()())))
  – each level of parentheses brackets the elements of a list
    • head and remainder of list
    • each node is denoted as a construct cell or atom
Pointers and Recursive Types
Pointers and Recursive Types

• references in functional languages are acyclic
  – circular references typically found only in imperative languages

• often need mutually recursive types
  – example: symbol table and syntax tree nodes need to refer to each other
  – if declared one at a time, not able to refer to each other
Pointers and Recursive Types

• mutually recursive types
  – ML example: symbol table and syntax tree nodes declared together

  ```ml
datatype sym_tab_rec = variable of ...
  | type of ...
  | ...
  | subroutine of {code : syn_tree_node, ...}
  and syn_tree_node = expression of ...
  | loop of ...
  | ...
  | subr_call of {subr : sym_tab_rec, ...};
  ```
Pointers and Recursive Types

• tree types in value model languages
  – Pascal
    
    type chr_tree_ptr = ^chr_tree;
    chr_tree = record
        left, right : chr_tree_ptr;
        val : char
    end;
  
  – Ada
    
    type chr_tree;
    type chr_tree_ptr is access chr_tree;
    type chr_tree is record
        left, right : chr_tree_ptr;
        val : character;
    end record;
Pointers and Recursive Types

• tree types in value model languages (cont.)
  – C
    ```
    struct chr_tree {
        struct chr_tree *left, *right;
        char val;
    };
    ```

• for mutually recursive types
  – Pascal: forward references
  – Ada and C: incomplete data types
Pointers and Recursive Types
Pointers and Recursive Types

• allocation of space from the heap
  – Pascal
    
    \texttt{new(my\_ptr);}
  – Ada
    
    \texttt{my\_ptr := new chr\_tree;}
  – C
    
    \texttt{my\_ptr = (struct chr\_tree *) malloc(sizeof(struct chr\_tree));}
  • returns void*
  – C++, Java, C#
    
    \texttt{my\_ptr = new chr\_tree( arg\_list );}
Pointers and Recursive Types

• accessing objects pointed to by references
  – Pascal
    
    my_ptr^.val := 'X';
  
  – C
    
    (*my_ptr).val = 'X';
  • or for structs
    
    my_ptr->val = 'X';

  – Ada
    
    • depends on context
Pointers and Recursive Types

• most languages blur the distinction between l-values and r-values by implicitly dereferencing variables on right-hand side of assignment
Pointers and Recursive Types

• C pointers and arrays

```c
int n;
int *a;    /* pointer to integer */
int b[10]; /* array of 10 integers */
```

– all of the following are valid

1. `a = b;`  /* make a point to the initial element of b */
2. `n = a[3];`
3. `n = *(a+3);`  /* equivalent to previous line */
4. `n = b[3];`
5. `n = *(b+3);`  /* equivalent to previous line */
Pointers and Recursive Types

• subscript operator [ ]

\[ E1[E2] \]

– equivalent to

\[ (((E1)+(E2))) \]
\[ (((E2)+(E1))) \]

– order does not matter

\[ A[3] \]
\[ 3[A] \]
Pointers and Recursive Types

• C pointers and arrays closely linked
  
  int *a == int a[]
  int **a == int *a[]

• some subtle differences
  – a declaration allocates an array if it specifies a size in the first dimension
  – otherwise it allocates a pointer
  
  int **a, int *a[]; // pointer to pointer to int
  int *a[n];       // n-element array of row pointers
  int a[n][m];    // 2-d array
Pointers and Recursive Types

- declaring an array requires all dimensions except the last
  - invalid
    
    int a[][];
    int (*a)[];
  
  - C declaration rule: read right as far as you can (subject to parentheses), then left, then out a level and repeat
    
    int *a[n];    // n-element array of pointers to integer
    int (*a)[n];  // pointer to n-element array of integers
Pointers and Recursive Types

• when a heap-allocated object is no longer live, space must be reclaimed
  – stack objects reclaimed automatically
  – heap space can be reclaimed automatically with garbage collection
Pointers and Recursive Types

• some languages require heap space to be reclaimed by the programmer
  – Pascal
    
    ```
    dispose(my_ptr);
    ```
  – C
    
    ```
    free(my_ptr);
    ```
  – C++
    
    ```
    delete my_ptr;
    ```
  – if not reclaimed, results in memory leaks
• dangling reference
  – live pointer that no longer points to a valid object
  – only in languages that have explicit deallocation
Pointers and Recursive Types

• dangling reference examples
  – subroutine returns pointer to local variable
  – active pointer to space that has been reclaimed
    • even if the pointer were changed to NULL, other pointers might still refer to this space
    • changes may read or write bits that are parts of other objects
Pointers and Recursive Types

• explicit reclaiming can be a burden on programmers and a source of memory leaks

• alternative: garbage collection
  – automatic reclaiming of heap space
  – required for functional languages
  – found in Java, Modula-3, C#
  – difficult to implement
  – convenient for programmers: no dangling pointers
  – can be slow
Pointers and Recursive Types

• several types of garbage collection
  – reference counts
  – mark-and-sweep
  – pointer reversal
  – stop-and-copy
  – others
Pointers and Recursive Types

• reference counts
  – one way to designate an object as garbage is when no pointers to it exist
  – place a reference count in each object that keeps track of the number of pointers to it
  – when object is created, reference count is set to 1
  – when pointers are assigned or re-assigned, count is incremented/decremented
  – when count reaches 0, space can be reclaimed
Pointers and Recursive Types

• reference counts (cont.)
  – reference count also decremented upon subroutine return if local pointers involved
  – system must run recursively, so that any pointers in reclaimed space are accounted for
  – pointers must otherwise be initialized to null
  – works well with strings, which have no circular dependencies
Pointers and Recursive Types

- the heap is a chain of nodes (the `free_list`)
- each node has a reference count (RC)
- for an assignment, like \( q = p \), garbage can occur
Pointers and Recursive Types

- reference counts may miss circular structures
Pointers and Recursive Types

• disadvantages
  - failure to detect inaccessible circular chains
  - storage overhead created by appending an integer reference count to every node in the heap
  - performance overhead whenever a pointer is assigned or a heap block is allocated or deallocated
Pointers and Recursive Types

• some garbage collection schemes use tracing
  – instead of counting the number or references to an object
  – find all good space by following valid pointers that are active
  – work recursively through the heap, starting from external pointers
Pointers and Recursive Types

• mark-and-sweep
  – 3 main steps executed when space is low
  1. all blocks in the heap marked with 0
  2. follows pointers outside the heap and recursively explores all heap space; each newly discovered block marked with 1
     • if already marked with 1, can stop following that chain
  3. final walk through heap to remove every block still marked with 0
• triggered by `q=new node()` and `free_list = null`.
• all accessible nodes are marked 1
Pointers and Recursive Types

- now `free_list` is restored and
- the assignment `q=new node()` can proceed
Pointers and Recursive Types

• advantages over reference counting
  – reclaims all garbage in the heap
  – only invoked when the heap is full

• issues with mark-and-sweep
  – system must know beginning and ending points of blocks
    • each block must contain a size
  – collector must be able to find pointers contained in each block
  – recursive nature requires stack space
  – can be time-intensive
Pointers and Recursive Types

• pointer reversal
  – to avoid large stack during recursive search, use pointers already in heap space to work back up the chain
  – such pointers oriented in the wrong direction
  – can be reversed to point backward in the “stack”
  – collector keeps track of current block and the block from whence it came
Pointers and Recursive Types

• heap exploration using pointer reversal
Pointers and Recursive Types

• stop-and-copy (aka copy collection)
  – heap space divided into two halves
    • only one is active at any given time
  – when one half nearly full, collector recursively explores heap, copying all valid blocks to other half
  – defragments while copying
  – pointers to blocks changed during copy
  – when finished, only useful blocks in heap; other half is no longer used
  – on next collection, halves are reversed
Pointers and Recursive Types

- triggered by `q=new node()` and `free_list` outside the active half:
Pointers and Recursive Types

• accessible nodes copied to other half
  – the accessible nodes are packed, orphans are returned to the free_list, and the two halves reverse roles
Pointers and Recursive Types

• issues with stop-and-copy
  – only half of the heap can be used at any one time
  – incurs additional overhead
  – can be time-intensive
Pointers and Recursive Types

• other techniques
  – generational collection
    • since most dynamically allocated blocks are short-lived, newer ones stored in half the heap
    • this half is checked first if garbage collection necessary
    • during garbage collection, valid blocks transferred to long-term heap
    • system must examine long-term heap if more space needed, and to reduce memory leaks
  – conservative collection
    • if blocks in the heap do not state explicitly the location of pointers, any space that “looks like” a pointer can be considered such and used with mark-and-sweep
    • may leave some useless blocks, but guaranteed not to delete currently used blocks
Pointers and Recursive Types

• summary
  • modern algorithms are more elaborate
    • most are hybrids/refinements of the above three
  • in Java, garbage collection is built-in
    • runs as a low-priority thread
    • also, `System.gc` may be called by the program
  • functional languages have garbage collection built-in
  • C/C++ default garbage collection to the programmer

Lists

• a list is defined recursively as either the empty list or a pair consisting of an object (which may be either a list or an atom) and another (shorter) list
  – ideally suited to programming in functional and logic languages, which rely on recursion
    • Lisp: a program is a list, and can extend itself at run time by constructing a list and executing it
  – lists can also be used in imperative programs
Lists

- list operations tend to generate garbage
  - useful to have automatic garbage collection
- some languages (e.g., ML) require lists to contain items of the same type
- other languages (e.g., Python, Lisp) allow lists to contain items of differing types
  - implemented as chain of two pointers: one to element and one to next block
Lists

• since everything in Lisp is a list, differentiation required between lists to be evaluated and constant lists

• Lisp list constructors

\[
\begin{align*}
(\text{cons }'a' '(b)) & \implies (a\ b) \\
(\text{car }'(a\ b)) & \implies a \\
(\text{car }\text{nil}) & \implies \text{null} \\
(\text{cdr }'(a\ b\ c)) & \implies (b\ c) \\
(\text{cdr }'(a)) & \implies \text{nil} \\
(\text{cdr }\text{nil}) & \implies \text{null} \\
(\text{append }'(a\ b) ' '(c\ d)) & \implies (a\ b\ c\ d)
\end{align*}
\]
Lists

• list comprehension (e.g., Haskell, Python)
  – Python example: all odd numbers less than 100
    \[ [i^2 \text{ for } i \text{ in range}(1, 100) \text{ if } i \mod 2 == 1] \]
Files and Input/Output

• input/output (I/O) facilities allow a program to communicate with the outside world
  – interactive I/O generally implies communication with human users or physical devices
  – files generally refer to off-line storage implemented by the operating system
• files may be further categorized into
  – temporary
  – persistent
Equality Testing and Assignment

- when an expression is encountered testing for equality, how should it be interpreted?
  - for primitive types, easily defined
  - example: $s = t$ for character strings
    - are $s$ and $t$ aliases?
    - occupy the same storage that is bit-wise identical over its full length?
    - contain the same sequence of characters?
    - would appear the same if printed?
  - shallow vs. deep comparison
Equality Testing and Assignment

• Scheme has distinct equality-testing functions for each

\[(\text{eq? } a \ b)\] ; do a and b refer to the same object?
\[(\text{eqv? } a \ b)\] ; are a and b provably semantically equivalent?
\[(\text{equal? } a \ b)\] ; do a and b have the same recursive structure?

– behavior

\[(\text{eq? } \#t \ \#t)\] \implies \#t (true)
\[(\text{eq? } \text{\textquoteleft}foo\textquoteleft\text{\textquoteleft}foo)\] \implies \#t
\[(\text{eq? } \text{\textquoteleft}(a \ b)\textquoteleft(a \ b))\] \implies \#f (false); created by separate cons-es
\[(\text{let } ((\text{p } (a \ b)))\]
\((\text{eq? } \text{p } \text{p}))\] \implies \#t; created by the same cons
\[(\text{eq? } 2 \ 2)\] \implies \text{unspecified}
\[(\text{eq? } \text{\textquoteleft}foo\textquoteleft\text{\textquoteleft}foo\textquoteleft)\] \implies \text{unspecified}

• deep assignments rare in any language