Designing, Coding, and Documenting

Lecture 10
CS211 – Summer 2008

Software is becoming very complicated…

Real systems are large, complex, buggy, bloated, unmaintainable.

Year	*Operating System*	*Millions of lines of code**
1993	Windows NT 3.1	6
1994	Windows NT 3.5	10
1996	Windows NT 4.0	16
2000	Windows 2000	29
2001	Windows XP	40
2005	Windows Vista Beta 2	50

Commercial software typically has 20 to 30 bugs for every 1,000 lines of code!!†

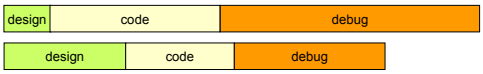
*source: Wikipedia
†source: CMU CyLab Sustainable Computing Consortium

Designing and Writing a Program

- Design stage – *THINK* about
  - the data you are working with
  - the operations you will perform
  - The data structures you will use to represent it
  - how to structure your program for abstraction and encapsulation
- Coding stage – code in small bits
  - test as you go
  - understand preconditions and postconditions
  - insert sanity checks (assert statements in Java are good)
  - worry about corner cases
- Use Java API to advantage

The Design-Code-Debug Cycle

- Design is faster than debugging (and more fun)
  - extra time spent designing reduces coding and debugging
- Which is better?



- Actually, should be more like this:

Divide and Conquer!

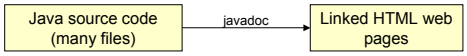
- Break program into small parts that can be implemented, tested in isolation
- Define interfaces for parts that talk to each other – develop *contracts* (preconditions, postconditions)
- Make sure contracts are obeyed
  - Clients use interfaces correctly
  - Implementers implement interfaces correctly (test!)
- Key: good interface documentation

Documentation is Code

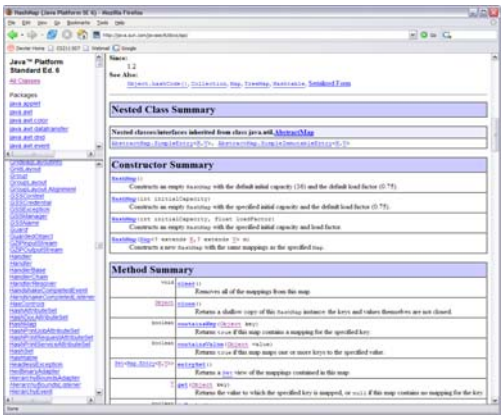
- Comments are as important as the code itself
  - determine successful use of code
  - determine whether code can be maintained
  - creation/maintenance = 1/10
- Documentation belongs in code
  - Otherwise as code evolves, documentation drifts away
  - Put specs in comments next to code when possible
  - Separate documentation? Code should link to it.
- Avoid useless comments
  - `x = x + 1; //add one to x`
  - Need to document algorithm? Write a paragraph at the top.
  - Or break method into smaller, clearer pieces.

Javadoc

- An important Java documentation tool



- Extracts documentation from classes, interfaces
  - Requires properly formatted comments
- Produces browsable, hyperlinked HTML web pages



How Javadoc is Produced

indicates Javadoc comment

Javadoc keywords

can include HTML

```
/**
 * Constructs an empty <tt>HashMap</tt> with the specified initial
 * capacity and the default load factor (0.75).
 *
 * @param  initialCapacity the initial capacity.
 * @throws IllegalArgumentException if the initial capacity is negative.
 */
public HashMap(int initialCapacity) {
    this(initialCapacity, DEFAULT_LOAD_FACTOR);
}

/**
 * Constructs an empty <tt>HashMap</tt> with the default initial capacity
 * (16) and the default load factor (0.75).
 */
public HashMap() {
    this.loadFactor = DEFAULT_LOAD_FACTOR;
    threshold = (int)(DEFAULT_INITIAL_CAPACITY * DEFAULT_LOAD_FACTOR);
    table = new Entry[DEFAULT_INITIAL_CAPACITY];
    init();
}
```


Some Useful Javadoc Tags

`@return` *description*
- Describes the return value of the method, if any
- E.g., **`@return the sum of the two intervals`**

`@param` *parameter-name description*
- Describes the parameters of the method
- E.g., **`@param i the other interval`**

`@author` *name*

`@deprecated` *reason*

`@see package.class#member`

`{@code` *expression*`}`
- Puts expression in code font

Developing and Documenting an ADT

1. Write an overview – purpose of the ADT
2. Decide on a set of supported operations
3. Write a specification for each operation

1. Writing an ADT Overview

- Example abstraction: a closed interval [a,b] on the real number line
  - [a,b] = { x | a ≤ x ≤ y }
- Example overview:

```
/**
 * An Interval represents a closed interval [a,b]
 * on the real number line.
 */
```

Javadoc comment

Abstract description of the ADT's values

2. Identify the Operations

- Enough operations for needed tasks
- Avoid unnecessary operations – keep it simple!
  - Don't include operations that client (without access to internals of class) can implement

3. Writing Method Specifications

- Include
  - Signature: types of method arguments, return type
  - Description of what the method does (abstractly)
- Good description (definitional)

```
/** Add two intervals. The sum of two intervals is
 * a set of values containing all possible sums of
 * two values, one from each of the two intervals.
 */
public Interval plus(Interval i);
```

- Bad description (operational)

```
/** Return a new Interval with lower bound a+i.a,
 * upper bound b+i.b.
 */
public Interval plus(Interval i);
```

Not abstract, might as well read the code…

3. Writing Specifications (cont'd)

- Attach before methods of class or interface

```
/** Add two intervals. The sum of two intervals is
 * a set of values containing all possible sums of
 * two values, one from each of the two intervals.
 *
 * @param i the other interval
 * @return the sum of the two intervals
 */
```

Method overview
Method description
Additional tagged clauses

Know Your Audience

- Code and specs have a target audience
  - the programmers who will maintain and use it
- Code and specs should be written…
  - with enough documented detail so they can understand it
  - while avoiding spelling out the obvious
- Try it out on the audience when possible
  - design reviews before coding
  - code reviews

Consistency

A foolish consistency is the hobgoblin of little minds.
– Emerson

- Pick a consistent coding style, stick with it
  - Make your code understandable by "little minds"
- Teams should set common style
- Match style when *editing* someone **else's** *code*
  - Not just syntax, also design style

Simplicity

The present letter is a very long one, simply because I had no time to make it shorter. –Blaise Pascal

Be brief. –Strunk & White

- Applies to programming. Simple code is:
  - Easier and quicker to understand
  - More likely to be correct
- Good code is simple, short, and clear
  - Use complex algorithms, data structures only when they are needed
  - Always reread code (and writing) to see if it can be made shorter, simpler, clearer

Choosing Names

- Don't try to document with variable names
  - Longer is not necessarily better

```
int searchForElement(
   int[] array_of_elements_to_search,
   int element_to_look_for);

int search(int[] array, int target);
```

- Names should be short but suggestive
- Local variable names should be short

Avoid Copy-and-Paste

- Biggest single source of program errors
  - Bug fixes never reach all the copies
  - Think twice before using your editor's copy-and-paste function

- Abstract instead of copying!
  - Write many calls to a single function rather than copying the same block of code around

Design vs Programming by Example

- Programming by example:
  - copy code that does something like what you want
  - hack it until it works
- Problems:
  - inherit bugs in code
  - don't understand code fully
  - usually inherit unwanted functionality
  - code is a bolted-together hodge-podge
- Alternative: design
  - understand exactly why your code works
  - reuse abstractions, not code templates

What Makes a Good Algorithm?

- Suppose you have two algorithms or data structures that basically do the same thing. Which is *better*?
  - Faster?
  - Less space?
  - Easier to code?
  - Easier to maintain?
  - Easier to understand?
- How do we measure time and space for an algorithm?

Inheritance vs Encapsulation

- Inheritance is useful, but it is possible to overdo it
  - Overused by many Java & OO programmers
- `class C extends D` means state and methods of D are accessible in C
  - Tempting, often useful, but also can be dangerous!
  - C becomes a subtype of D
- Inherit only if a C should be used as a D
  - all methods of D should still make sense
  - A function expecting a D will work on a C
- Prefer Java interfaces instead

Avoid Premature Optimization

Premature optimization is the root of all evil (or at least most of it) in programming. --Donald Knuth

- Temptations to avoid
  - Copying code to avoid overhead of abstraction mechanisms
  - Using complex algorithms & data structures unnecessarily
  - Violating abstraction barriers
- Result:
  - Less simple and clear
  - Performance gains often negligible
- Avoid trying to accelerate performance until
  - You have the program designed and working
  - You know that simplicity needs to be sacrificed
  - You know where simplicity needs to be sacrificed

Avoid Duplication

- Duplication in source code creates an implicit constraint to maintain, a quick path to failure
  - Duplicating code fragments (by copying)
  - Duplicating specs in classes and in interfaces
  - Duplicating specifications in code and in external documents
  - Duplicating same information on many web pages
- Solutions:
  - Named abstractions (e.g., declaring methods)
  - Indirection (linking pointers)
  - Generate duplicate information from source (e.g., Javadoc!)
- *If you must duplicate*:
  - Make duplicates link to each other so always can find all clones

Pair Programming

- Pilot/copilot
  - pilot codes, copilot watches and makes suggestions
  - pilot must convince copilot that code works
  - take turns
- Or: work independently on different parts after deciding on an interface
  - frequent design review
  - each programmer must convince the other
  - reduces debugging time
- Test everything

How to Make your Group Project Harder

1. Have one person do all the work. That person burns out, and no one else can finish the project.
2. Decide that the other member(s) of your group are useless. Don't communicate or meet with them.
3. Decide that all the other members of your group are useless. You are the lone master hacker. Charge off and code everything without talking to anyone else. Unless you are very unlucky, you'll make some bad assumption that forces all your code to be thrown out anyway.

How to Make your Group Project Harder

4. Everyone implements pieces of the system with no discussion of how they will fit together until just before the assignment is due. You won't be able to glue it all together in time.
5. Work extremely closely all the time, spending all your time talking rather than doing actual implementation; the group will slow down to the speed of the slowest person.
6. Don't start until three days before the assignment is due. Pull three all-nighters in a row. With lack of sleep you will write broken code. With luck, you will get sick, miss some other classes as well.

How to Make your Group Project Harder

7. Don't ask the TAs or the instructors any questions when design problems come up; put off working on the project and hope the problems will magically solve themselves.
8. Don't use any of the techniques for software design that you learn in this class. This works best if you don't attend class at all – avoid polluting your mind.

Team programming

- A large project may have dozens, hundreds, thousands of programmers all working at once
  - E.g. Windows Vista: >2,000 programmers, >50M lines of code
  - Great for efficiency
  - But very difficult to manage
- How do you manage huge software projects?
  - High-level design (e.g well-defined interfaces) becomes key
  - Formal software engineering methods

Source control

- Multiple people editing the same files at once => disaster!
- Source control (or revision control) software helps manage multiple versions of source code files
  - Popular packages: rcs, cvs, svn, SourceSafe, ClearCase
- Idea: code is maintained in some central repository
  - To work on a file, you *check out* the file from the repository
  - Edit the file on your local computer
  - When done, you *commit* the file to the repository
  - Source control maintains a historical *log* of all changes

Debugging

- Debugging usually requires more time and just as much talent as writing new code
  - Try to isolate the problem
    - Add println statements to output temporary variables.
    - Use binary search.
    - Temporarily remove or comment out unrelated code. (But keep a copy of your original version!)
  - Take a break
    - Think about what could be causing the observed behavior
  - Use an IDE with a debugger

Debugging

- Don't just randomly change code. It never works.
  - It just frustrates the programmer even more!
  - Better to take a step back and re-think your code at a conceptual level.
- Mysterious compiler or run-time errors
  - Tip: try google-ing the error. You'll usually find information about the error.
- Common beginning programmer excuse: "It must be a bug in the compiler!"
  - The chances of this are incredibly small.

No Silver Bullets

- These are all rules of thumb; but there is no panacea, and every rule has its exceptions
- You can only learn by doing – we can't do it for you
- Following software engineering rules only makes success more likely!