The Objective-C Programming Language

Tools & Languages: Objective-C

2009-10-19

Apple Inc.
1 Infinite Loop
Cupertino, CA 95014
408-996-1010

Simultaneously published in the United States and Canada.

Contents

Figures and Listings

Introduction to The Objective-C Programming Language

The Objective-C language is a simple computer language designed to enable sophisticated object-oriented programming. Objective-C is defined as a small but powerful set of extensions to the standard ANSI C language. Its additions to C are mostly based on Smalltalk, one of the first object-oriented programming languages. Objective-C is designed to give C full object-oriented programming capabilities, and to do so in a simple and straightforward way.

Most object-oriented development environments consist of several parts:

- An object-oriented programming language
- A library of objects
- A suite of development tools
- A runtime environment

This document is about the first component of the development environment—the programming language. It fully describes the Objective-C language, and provides a foundation for learning about the second component, the Mac OS X Objective-C application frameworks—collectively known as Cocoa. You can start to learn more about Cocoa by reading *Getting Started with Cocoa*. The two main development tools you use are Xcode and Interface Builder, described in *Xcode Workspace Guide* and *Interface Builder* respectively. The runtime environment is described in a separate document, *Objective-C Runtime Programming Guide*.

Important: This document describes the version of the Objective-C language released in Mac OS X v10.6, which introduces the associative references feature (see "Associative References" (page 83)). To learn about version 1.0 of the Objective-C language (available in in Mac OS X v10.4 and earlier), read *Object Oriented Programming and the Objective-C Programming Language 1.0*.

Who Should Read This Document

The document is intended for readers who might be interested in:

- Programming in Objective-C
- Finding out about the basis for the Cocoa application framework

This document both introduces the object-oriented model that Objective-C is based upon and fully documents the language. It concentrates on the Objective-C extensions to C, not on the C language itself.

Because this isn't a document about C, it assumes some prior acquaintance with that language. However, it doesn't have to be an extensive acquaintance. Object-oriented programming in Objective-C is sufficiently different from procedural programming in ANSI C that you won't be hampered if you're not an experienced C programmer.

Organization of This Document

This document is divided into several chapters and one appendix.

The following chapters cover all the features Objective-C adds to standard C.

- "Objects, Classes, and Messaging" (page 13)
- "Defining a Class" (page 35)
- "Allocating and Initializing Objects" (page 47)
- "Protocols" (page 57)
- "Declared Properties" (page 67)
- "Categories and Extensions" (page 79)
- "Associative References" (page 83)
- "Fast Enumeration" (page 87)
- "Enabling Static Behavior" (page 91)
- "Selectors" (page 95)
- "Exception Handling" (page 99)
- "Threading" (page 103)
- "Remote Messaging" (page 105)

The Apple compilers are based on the compilers of the GNU Compiler Collection. Objective-C syntax is a superset of GNU C/C++ syntax, and the Objective-C compiler works for C, C++ and Objective-C source code. The compiler recognizes Objective-C source files by the filename extension `.m`, just as it recognizes files containing only standard C syntax by filename extension `.c`. Similarly, the compiler recognizes C++ files that use Objective-C by the extension `.mm`. Other issues when using Objective-C with C++ are covered in "Using C++ With Objective-C" (page 111).

The appendix contains reference material that might be useful for understanding the language:

- "Language Summary" (page 117) lists and briefly comments on all of the Objective-C extensions to the C language.

Conventions

Where this document discusses functions, methods, and other programming elements, it makes special use of computer voice and italic fonts. Computer voice denotes words or characters that are to be taken literally (typed as they appear). Italic denotes words that represent something else or can be varied. For example, the syntax:

`@interface`*ClassName*`(`*CategoryName*`)`

means that `@interface` and the two parentheses are required, but that you can choose the class name and category name.

Where example code is shown, ellipsis points indicates the parts, often substantial parts, that have been omitted:

```
- (void)encodeWithCoder:(NSCoder *)coder
{
    [super encodeWithCoder:coder];
    ...
}
```

The conventions used in the reference appendix are described in that appendix.

See Also

If you have never used object-oriented programming to create applications before, you should read *Object-Oriented Programming with Objective-C*. You should also consider reading it if you have used other object-oriented development environments such as C++ and Java, since those have many different expectations and conventions from Objective-C. *Object-Oriented Programming with Objective-C* is designed to help you become familiar with object-oriented development from the perspective of an Objective-C developer. It spells out some of the implications of object-oriented design and gives you a flavor of what writing an object-oriented program is really like.

Runtime

Objective-C Runtime Programming Guide describes aspects of the Objective-C runtime and how you can use it.

Objective-C Runtime Reference describes the data structures and functions of the Objective-C runtime support library. Your programs can use these interfaces to interact with the Objective-C runtime system. For example, you can add classes or methods, or obtain a list of all class definitions for loaded classes.

Objective-C Release Notes describes some of the changes in the Objective-C runtime in the latest release of Mac OS X.

Memory Management

Objective-C supports two environments for memory management: automatic garbage collection and reference counting:

- *Garbage Collection Programming Guide* describes the garbage collection system used by Cocoa. (Not available on iPhone—you cannot access this document through the iPhone Dev Center.)
- *Memory Management Programming Guide* describes the reference counting system used by Cocoa.

Objects, Classes, and Messaging

This chapter describes the fundamentals of objects, classes, and messaging as used and implemented by the Objective-C language. It also introduces the Objective-C runtime.

Runtime

The Objective-C language defers as many decisions as it can from compile time and link time to runtime. Whenever possible, it dynamically performs operations such as creating objects and determining what method to invoke. This means that the language requires not just a compiler, but also a runtime system to execute the compiled code. The runtime system acts as a kind of operating system for the Objective-C language; it's what makes the language work. Typically, however, you don't need to interact with the runtime directly. To understand more about the functionality it offers, though, see *Objective-C Runtime Programming Guide*.

Objects

As the name implies, object-oriented programs are built around **objects**. An object associates data with the particular operations that can use or affect that data. Objective-C provides a data type to identify an object variable without specifying a particular class of the object—this allows for dynamic typing. In a program, you should typically ensure that you dispose of objects that are no longer needed.

Object Basics

An object associates data with the particular operations that can use or affect that data. In Objective-C, these operations are known as the object's **methods**; the data they affect are its **instance variables**. In essence, an object bundles a data structure (instance variables) and a group of procedures (methods) into a self-contained programming unit.

For example, if you are writing a drawing program that allows a user to create images composed of lines, circles, rectangles, text, bit-mapped images, and so forth, you might create classes for many of the basic shapes that a user can manipulate. A Rectangle object, for instance, might have instance variables that identify the position of the rectangle within the drawing along with its width and its height. Other instance variables could define the rectangle's color, whether or not it is to be filled, and a line pattern that should be used to display the rectangle. A Rectangle class would have methods to set an instance's position, size, color, fill status, and line pattern, along with a method that causes the instance to display itself.

In Objective-C, an object's instance variables are internal to the object; generally, you get access to an object's state only through the object's methods (you can specify whether subclasses or other objects can access instance variables directly by using scope directives, see "The Scope of Instance Variables" (page 40)). For others to find out something about an object, there has to be a method to supply the information. For example, a Rectangle would have methods that reveal its size and its position.

Moreover, an object sees only the methods that were designed for it; it can't mistakenly perform methods intended for other types of objects. Just as a C function protects its local variables, hiding them from the rest of the program, an object hides both its instance variables and its method implementations.

id

In Objective-C, object identifiers are a distinct data type: `id`. This is the general type for any kind of object regardless of class. (It can be used for both instances of a class and class objects themselves.) `id` is defined as pointer to an object data structure:

```
typedef struct objc_object {
    Class isa;
} *id;
```

All objects thus have an `isa` variable that tells them of what class they are an instance.

Terminology: Since the Class type is itself defined as a pointer:

```
typedef struct objc_class *Class;
```

the `isa` variable is frequently referred to as the "`isa` pointer."

Like a C function or an array, an object is therefore identified by its address. All objects, regardless of their instance variables or methods, are of type `id`.

```
id anObject;
```

For the object-oriented constructs of Objective-C, such as method return values, `id` replaces `int` as the default data type. (For strictly C constructs, such as function return values, `int` remains the default type.)

The keyword `nil` is defined as a null object, an `id` with a value of `0`. `id`, `nil`, and the other basic types of Objective-C are defined in the header file `objc/objc.h`.

Dynamic Typing

The `id` type is completely nonrestrictive. By itself, it yields no information about an object, except that it is an object.

But objects aren't all the same. A Rectangle won't have the same methods or instance variables as an object that represents a bit-mapped image. At some point, a program needs to find more specific information about the objects it contains—what the object's instance variables are, what methods it can perform, and so on. Since the `id` type designator can't supply this information to the compiler, each object has to be able to supply it at runtime.

The `isa` instance variable identifies the object's **class**—what kind of object it is. Every Rectangle object would be able to tell the runtime system that it is a Rectangle. Every Circle can say that it is a Circle. Objects with the same behavior (methods) and the same kinds of data (instance variables) are members of the same class.

Objects are thus **dynamically typed** at runtime. Whenever it needs to, the runtime system can find the exact class that an object belongs to, just by asking the object. (To learn more about the runtime, see *Objective-C Runtime Programming Guide*.) Dynamic typing in Objective-C serves as the foundation for dynamic binding, discussed later.

The `isa` variable also enables objects to perform **introspection**—to find out about themselves (or other objects). The compiler records information about class definitions in data structures for the runtime system to use. The functions of the runtime system use `isa`, to find this information at runtime. Using the runtime system, you can, for example, determine whether an object implements a particular method, or discover the name of its superclass.

Object classes are discussed in more detail under "Classes" (page 23).

It's also possible to give the compiler information about the class of an object by statically typing it in source code using the class name. Classes are particular kinds of objects, and the class name can serve as a type name. See "Class Types" (page 27) and "Enabling Static Behavior" (page 91).

Memory Management

In an Objective-C program, it is important to ensure that objects are deallocated when they are no longer needed—otherwise your application's memory footprint becomes larger than necessary. It is also important to ensure that you do not deallocate objects while they're still being used.

Objective-C offers two environments for memory management that allow you to meet these goals:

- **Reference counting**, where you are ultimately responsible for determining the lifetime of objects.

  Reference counting is described in *Memory Management Programming Guide*.

- **Garbage collection**, where you pass responsibility for determining the lifetime of objects to an automatic "collector."

  Garbage collection is described in *Garbage Collection Programming Guide*. (Not available on iPhone—you cannot access this document through the iPhone Dev Center.)

Object Messaging

This section explains the syntax of sending messages, including how you can nest message expressions. It also discusses the "visibility" of an object's instance variables, and the concepts of polymorphism and dynamic binding.

Message Syntax

To get an object to do something, you send it a **message** telling it to apply a method. In Objective-C, **message expressions** are enclosed in brackets:

```
[receiver message]
```

The receiver is an object, and the message tells it what to do. In source code, the message is simply the name of a method and any arguments that are passed to it. When a message is sent, the runtime system selects the appropriate method from the receiver's repertoire and invokes it.

For example, this message tells the `myRectangle` object to perform its `display` method, which causes the rectangle to display itself:

```
[myRectangle display];
```

The message is followed by a " ; " as is normal for any line of code in C.

The method name in a message serves to "select" a method implementation. For this reason, method names in messages are often referred to as **selectors**.

Methods can also take parameters, or "arguments." A message with a single argument affixes a colon (`:`) to the selector name and puts the argument right after the colon. This construct is called a keyword; a keyword ends with a colon, and an argument follows the colon, as shown in this example:

```
[myRectangle setWidth:20.0];
```

A selector name includes all keywords, including colons, but does not include anything else, such as return type or parameter types. The imaginary message below tells the `myRectangle` object to set its origin to the coordinates (30.0, 50.0):

```
[myRectangle setOrigin:30.0 :50.0]; // This is a bad example of multiple arguments
```

Since the colons are part of the method name, the method is named `setOrigin::`. It has two colons as it takes two arguments. This particular method does not interleave the method name with the arguments and, thus, the second argument is effectively unlabeled and it is difficult to determine the kind or purpose of the method's arguments.

Instead, method names should interleave the name with the arguments such that the method's name naturally describes the arguments expected by the method. For example, the Rectangle class could instead implement a `setOriginX:y:` method that makes the purpose of its two arguments clear:

```
[myRectangle setOriginX: 30.0 y: 50.0]; // This is a good example of multiple
arguments
```

Important: The sub-parts of the method name—of the selector—are not optional, nor can their order be varied. "Named arguments" and "keyword arguments" often carry the implication that the arguments to a method can vary at runtime, can have default values, can be in a different order, can possibly have additional named arguments. This is not the case with Objective-C.

For all intents and purposes, an Objective-C method declaration is simply a C function that prepends two additional arguments (see Messaging in the *Objective-C Runtime Programming Guide*). This is different from the named or keyword arguments available in a language like Python:

```
def func(a, b, NeatMode=SuperNeat, Thing=DefaultThing):
    pass
```

where Thing (and NeatMode) might be omitted or might have different values when called.

Methods that take a variable number of arguments are also possible, though they're somewhat rare. Extra arguments are separated by commas after the end of the method name. (Unlike colons, the commas aren't considered part of the name.) In the following example, the imaginary `makeGroup:` method is passed one required argument (**group**) and three that are optional:

```
[receiver makeGroup:group, memberOne, memberTwo, memberThree];
```

Like standard C functions, methods can return values. The following example sets the variable `isFilled` to `YES` if `myRectangle` is drawn as a solid rectangle, or `NO` if it's drawn in outline form only.

```
BOOL isFilled;
isFilled = [myRectangle isFilled];
```

Note that a variable and a method can have the same name.

One message expression can be nested inside another. Here, the color of one rectangle is set to the color of another:

```
[myRectangle setPrimaryColor:[otherRect primaryColor]];
```

Objective-C also provides a dot (`.`) operator that offers a compact and convenient syntax for invoking an object's accessor methods. This is typically used in conjunction with the declared properties feature (see "Declared Properties" (page 67)), and is described in "Dot Syntax" (page 20).

Sending Messages to nil

In Objective-C, it is valid to send a message to `nil`—it simply has no effect at runtime. There are several patterns in Cocoa that take advantage of this fact. The value returned from a message to `nil` may also be valid:

- If the method returns an object, then a message sent to `nil` returns `0` (`nil`), for example:

  ```
  Person *motherInLaw = [[aPerson spouse] mother];
  ```

  If `aPerson`'s spouse is `nil`, then `mother` is sent to `nil` and the method returns `nil`.
- If the method returns any pointer type, any integer scalar of size less than or equal to `sizeof(void*)`, a `float`, a `double`, a `long double`, or a `long long`, then a message sent to `nil` returns `0`.

- If the method returns a struct, as defined by the *Mac OS X ABI Function Call Guide* to be returned in registers, then a message sent to `nil` returns `0.0` for every field in the data structure. Other struct data types will not be filled with zeros.
- If the method returns anything other than the aforementioned value types the return value of a message sent to `nil` is undefined.

The following code fragment illustrates valid use of sending a message to `nil`.

```
id anObjectMaybeNil = nil;

// this is valid
if ([anObjectMaybeNil methodThatReturnsADouble] == 0.0)
{
    // implementation continues...
}
```

Note: The behavior of sending messages to nil changed slightly with Mac OS X v10.5.

On Mac OS X v10.4 and earlier, a message to `nil` also is valid, as long as the message returns an object, any pointer type, `void`, or any integer scalar of size less than or equal to `sizeof(void*)`; if it does, a message sent to `nil` returns `nil`. If the message sent to `nil` returns anything other than the aforementioned value types (for example, if it returns any struct type, any floating-point type, or any vector type) the return value is undefined. You should therefore not rely on the return value of messages sent to `nil` unless the method's return type is an object, any pointer type, or any integer scalar of size less than or equal to `sizeof(void*)`.

The Receiver's Instance Variables

A method has automatic access to the receiving object's instance variables. You don't need to pass them to the method as arguments. For example, the `primaryColor` method illustrated above takes no arguments, yet it can find the primary color for `otherRect` and return it. Every method assumes the receiver and its instance variables, without having to declare them as arguments.

This convention simplifies Objective-C source code. It also supports the way object-oriented programmers think about objects and messages. Messages are sent to receivers much as letters are delivered to your home. Message arguments bring information from the outside to the receiver; they don't need to bring the receiver to itself.

A method has automatic access only to the receiver's instance variables. If it requires information about a variable stored in another object, it must send a message to the object asking it to reveal the contents of the variable. The `primaryColor` and `isFilled` methods shown above are used for just this purpose.

See "Defining a Class" (page 35) for more information on referring to instance variables.

Polymorphism

As the examples above illustrate, messages in Objective-C appear in the same syntactic positions as function calls in standard C. But, because methods "belong to" an object, messages behave differently than function calls.

In particular, an object can be operated on by only those methods that were defined for it. It can't confuse them with methods defined for other kinds of object, even if another object has a method with the same name. This means that two objects can respond differently to the same message. For example, each kind of object sent a `display` message could display itself in a unique way. A Circle and a Rectangle would respond differently to identical instructions to track the cursor.

This feature, referred to as **polymorphism**, plays a significant role in the design of object-oriented programs. Together with dynamic binding, it permits you to write code that might apply to any number of different kinds of objects, without you having to choose at the time you write the code what kinds of objects they might be. They might even be objects that will be developed later, by other programmers working on other projects. If you write code that sends a `display` message to an `id` variable, any object that has a `display` method is a potential receiver.

Dynamic Binding

A crucial difference between function calls and messages is that a function and its arguments are joined together in the compiled code, but a message and a receiving object aren't united until the program is running and the message is sent. Therefore, the exact method that's invoked to respond to a message can only be determined at runtime, not when the code is compiled.

The precise method that a message invokes depends on the receiver. Different receivers may have different method implementations for the same method name (polymorphism). For the compiler to find the right method implementation for a message, it would have to know what kind of object the receiver is—what class it belongs to. This is information the receiver is able to reveal at runtime when it receives a message (dynamic typing), but it's not available from the type declarations found in source code.

The selection of a method implementation happens at runtime. When a message is sent, a runtime messaging routine looks at the receiver and at the method named in the message. It locates the receiver's implementation of a method matching the name, "calls" the method, and passes it a pointer to the receiver's instance variables. (For more on this routine, see Messaging in the *Objective-C Runtime Programming Guide*.)

This **dynamic binding** of methods to messages works hand-in-hand with polymorphism to give object-oriented programming much of its flexibility and power. Since each object can have its own version of a method, a program can achieve a variety of results, not by varying the message itself, but by varying just the object that receives the message. This can be done as the program runs; receivers can be decided "on the fly" and can be made dependent on external factors such as user actions.

When executing code based upon the Application Kit, for example, users determine which objects receive messages from menu commands like Cut, Copy, and Paste. The message goes to whatever object controls the current selection. An object that displays text would react to a `copy` message differently from an object that displays scanned images. An object that represents a set of shapes would respond differently from a Rectangle. Since messages don't select methods (methods aren't bound to messages) until runtime, these differences are isolated in the methods that respond to the message. The code that sends the message doesn't have to be concerned with them; it doesn't even have to enumerate the possibilities. Each application can invent its own objects that respond in their own way to `copy` messages.

Objective-C takes dynamic binding one step further and allows even the message that's sent (the method selector) to be a variable that's determined at runtime. This is discussed in the section Messaging in the *Objective-C Runtime Programming Guide*.

Dynamic Method Resolution

You can provide implementations of class and instance methods at runtime using dynamic method resolution. See Dynamic Method Resolution in the *Objective-C Runtime Programming Guide* for more details.

Dot Syntax

Objective-C provides a dot (`.`) operator that offers a compact and convenient syntax you can use as an alternative to square bracket notation (`[]`s) to invoke accessor methods. It is particularly useful when you want to access or modify a property that is a property of another object (that is a property of another object, and so on).

Using the Dot Syntax

Overview

You can use the **dot syntax** to invoke accessor methods using the same pattern as accessing structure elements as illustrated in the following example:

```
myInstance.value = 10;
printf("myInstance value: %d", myInstance.value);
```

The dot syntax is purely "syntactic sugar"—it is transformed by the compiler into invocation of accessor methods (so you are not actually accessing an instance variable directly). The code example above is exactly equivalent to the following:

```
[myInstance setValue:10];
printf("myInstance value: %d", [myInstance value]);
```

General Use

You can read and write properties using the dot (`.`) operator, as illustrated in the following example.

Listing 1-1 Accessing properties using the dot syntax

```
Graphic *graphic = [[Graphic alloc] init];

NSColor *color = graphic.color;
CGFloat xLoc = graphic.xLoc;
BOOL hidden = graphic.hidden;
int textCharacterLength = graphic.text.length;

if (graphic.textHidden != YES) {
    graphic.text = @"Hello";
}
graphic.bounds = NSMakeRect(10.0, 10.0, 20.0, 120.0);
```

(`@"Hello"` is a constant `NSString` object—see "Compiler Directives" (page 118).)

Accessing a property *property* calls the get method associated with the property (by default, *property*) and setting it calls the set method associated with the property (by default, `set`*Property*`:`). You can change the methods that are invoked by using the Declared Properties feature (see "Declared Properties" (page 67)). Despite appearances to the contrary, the dot syntax therefore preserves encapsulation—you are not accessing an instance variable directly.

The following statements compile to exactly the same code as the statements shown in Listing 1-1 (page 20), but use square bracket syntax:

Listing 1-2 Accessing properties using bracket syntax

```
Graphic *graphic = [[Graphic alloc] init];

NSColor *color = [graphic color];
CGFloat xLoc = [graphic xLoc];
BOOL hidden = [graphic hidden];
int textCharacterLength = [[graphic text] length];

if ([graphic isTextHidden] != YES) {
    [graphic setText:@"Hello"];
}
[graphic setBounds:NSMakeRect(10.0, 10.0, 20.0, 120.0)];
```

An advantage of the dot syntax is that the compiler can signal an error when it detects a write to a read-only property, whereas at best it can only generate an undeclared method warning that you invoked a non-existent `set`*Property*`:` method, which will fail at runtime.

For properties of the appropriate C language type, the meaning of compound assignments is well-defined. For example, you could update the length property of an instance of `NSMutableData` using compound assignments:

```
NSMutableData *data = [NSMutableData dataWithLength:1024];
data.length += 1024;
data.length *= 2;
data.length /= 4;
```

which is equivalent to:

```
[data setLength:[data length] + 1024];
[data setLength:[data length] * 2];
[data setLength:[data length] / 4];
```

There is one case where properties cannot be used. Consider the following code fragment:

```
id y;
x = y.z;  // z is an undeclared property
```

Note that `y` is untyped and the `z` property is undeclared. There are several ways in which this could be interpreted. Since this is ambiguous, the statement is treated as an undeclared property error. If `z` is declared, then it is not ambiguous if there's only one declaration of a `z` property in the current compilation unit. If there are multiple declarations of a `z` property, as long as they all have the same type (such as `BOOL`) then it is legal. One source of ambiguity would also arise from one of them being declared `readonly`.

nil Values

If a `nil` value is encountered during property traversal, the result is the same as sending the equivalent message to `nil`. For example, the following pairs are all equivalent:

```
// each member of the path is an object
x = person.address.street.name;
x = [[[person address] street] name];

// the path contains a C struct
// will crash if window is nil or -contentView returns nil
y = window.contentView.bounds.origin.y;
y = [[window contentView] bounds].origin.y;

// an example of using a setter....
person.address.street.name = @"Oxford Road";
[[[person address] street] setName: @"Oxford Road"];
```

self

If you want to access a property of `self` using accessor methods, you must explicitly call out `self` as illustrated in this example:

```
self.age = 10;
```

If you do not use `self.`, you access the instance variable directly. In the following example, the set accessor method for the `age` property is *not* invoked:

```
age = 10;
```

Performance and Threading

The dot syntax generates code equivalent to the standard method invocation syntax. As a result, code using the dot syntax performs exactly the same as code written directly using the accessor methods. Since the dot syntax simply invokes methods, no additional thread dependencies are introduced as a result of its use.

Usage Summary

```
aVariable = anObject.aProperty;
```

Invokes the `aProperty` method and assigns the return value to `aVariable`. The type of the property `aProperty` and the type of `aVariable` must be compatible, otherwise you get a compiler warning.

```
anObject.name = @"New Name";
```

Invokes the `setName:` method on `anObject`, passing `@"New Name"` as the argument.

You get a compiler warning if `setName:` does not exist, if the property `name` does not exist, or if `setName:` returns anything but `void`.

```
xOrigin = aView.bounds.origin.x;
```

Invokes the `bounds` method and assigns `xOrigin` to be the value of the `origin.x` structure element of the `NSRect` returned by `bounds`.

```
NSInteger i = 10;
anObject.integerProperty = anotherObject.floatProperty = ++i;
```

Assigns `11` to both `anObject.integerProperty` and `anotherObject.floatProperty`. That is, the right hand side of the assignment is pre-evaluated and the result is passed to `setIntegerProperty:` and `setFloatProperty:`. The pre-evaluated result is coerced as required at each point of assignment.

Incorrect Use

The following patterns are strongly discouraged.

```
anObject.retain;
```

Generates a compiler warning (`warning: value returned from property not used.`).

```
/* method declaration */
- (BOOL) setFooIfYouCan: (MyClass *)newFoo;

/* code fragment */
anObject.fooIfYouCan = myInstance;
```

Generates a compiler warning that `setFooIfYouCan:` does not appear to be a setter method because it does not return `(void)`.

```
flag = aView.lockFocusIfCanDraw;
```

Invokes `lockFocusIfCanDraw` and assigns the return value to `flag`. This does not generate a compiler warning unless `flag`'s type mismatches the method's return type.

```
/* property declaration */
@property(readonly) NSInteger readonlyProperty;
/* method declaration */
- (void) setReadonlyProperty: (NSInteger)newValue;

/* code fragment */
self.readonlyProperty = 5;
```

Since the property is declared `readonly`, this code generates a compiler warning (`warning: assignment to readonly property 'readonlyProperty'`). Because the setter method is present, it will work at runtime, but simply adding a setter for a property does not imply `readwrite`.

Classes

An object-oriented program is typically built from a variety of objects. A program based on the Cocoa frameworks might use `NSMatrix` objects, `NSWindow` objects, `NSDictionary` objects, `NSFont` objects, `NSText` objects, and many others. Programs often use more than one object of the same kind or class—several `NSArray` objects or `NSWindow` objects, for example.

In Objective-C, you define objects by defining their class. The class definition is a prototype for a kind of object; it declares the instance variables that become part of every member of the class, and it defines a set of methods that all objects in the class can use.

The compiler creates just one accessible object for each class, a **class object** that knows how to build new objects belonging to the class. (For this reason it's traditionally called a "factory object.") The class object is the compiled version of the class; the objects it builds are **instances** of the class. The objects that do the main work of your program are instances created by the class object at runtime.

All instances of a class have the same set of methods, and they all have a set of instance variables cut from the same mold. Each object gets its own instance variables, but the methods are shared.

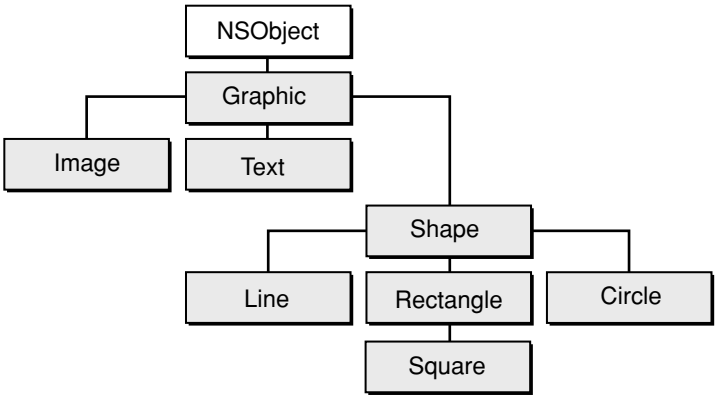
By convention, class names begin with an uppercase letter (such as "Rectangle"); the names of instances typically begin with a lowercase letter (such as "myRectangle").

Inheritance

Class definitions are additive; each new class that you define is based on another class from which it **inherits** methods and instance variables. The new class simply adds to or modifies what it inherits. It doesn't need to duplicate inherited code.

Inheritance links all classes together in a hierarchical tree with a single class at its root. When writing code that is based upon the Foundation framework, that root class is typically `NSObject`. Every class (except a root class) has a **superclass** one step nearer the root, and any class (including a root class) can be the superclass for any number of **subclasses** one step farther from the root. Figure 1-1 illustrates the hierarchy for a few of the classes used in the drawing program.

Figure 1-1 Some Drawing Program Classes



This figure shows that the Square class is a subclass of the Rectangle class, the Rectangle class is a subclass of Shape, Shape is a subclass of Graphic, and Graphic is a subclass of `NSObject`. Inheritance is cumulative. So a Square object has the methods and instance variables defined for Rectangle, Shape, Graphic, and `NSObject`, as well as those defined specifically for Square. This is simply to say that a Square object isn't only a Square, it's also a Rectangle, a Shape, a Graphic, and an `NSObject`.

Every class but `NSObject` can thus be seen as a specialization or an adaptation of another class. Each successive subclass further modifies the cumulative total of what's inherited. The Square class defines only the minimum needed to turn a Rectangle into a Square.

When you define a class, you link it to the hierarchy by declaring its superclass; every class you create must be the subclass of another class (unless you define a new root class). Plenty of potential superclasses are available. Cocoa includes the `NSObject` class and several frameworks containing definitions for more than 250 additional classes. Some are classes that you can use "off the shelf"—incorporate into your program as is. Others you might want to adapt to your own needs by defining a subclass.

Some framework classes define almost everything you need, but leave some specifics to be implemented in a subclass. You can thus create very sophisticated objects by writing only a small amount of code, and reusing work done by the programmers of the framework.

The NSObject Class

`NSObject` is a root class, and so doesn't have a superclass. It defines the basic framework for Objective-C objects and object interactions. It imparts to the classes and instances of classes that inherit from it the ability to behave as objects and cooperate with the runtime system.

A class that doesn't need to inherit any special behavior from another class should nevertheless be made a subclass of the `NSObject` class. Instances of the class must at least have the ability to behave like Objective-C objects at runtime. Inheriting this ability from the `NSObject` class is much simpler and much more reliable than reinventing it in a new class definition.

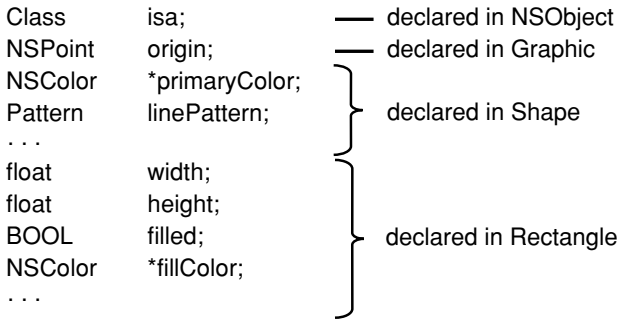
Note: Implementing a new root class is a delicate task and one with many hidden hazards. The class must duplicate much of what the `NSObject` class does, such as allocate instances, connect them to their class, and identify them to the runtime system. For this reason, you should generally use the `NSObject` class provided with Cocoa as the root class. For more information, see the Foundation framework documentation for the `NSObject` class and the `NSObject` protocol.

Inheriting Instance Variables

When a class object creates a new instance, the new object contains not only the instance variables that were defined for its class but also the instance variables defined for its superclass and for its superclass's superclass, all the way back to the root class. Thus, the `isa` instance variable defined in the `NSObject` class becomes part of every object. `isa` connects each object to its class.

Figure 1-2 shows some of the instance variables that could be defined for a particular implementation of Rectangle, and where they may come from. Note that the variables that make the object a Rectangle are added to the ones that make it a Shape, and the ones that make it a Shape are added to the ones that make it a Graphic, and so on.

Figure 1-2 Rectangle Instance Variables



A class doesn't have to declare instance variables. It can simply define new methods and rely on the instance variables it inherits, if it needs any instance variables at all. For example, Square might not declare any new instance variables of its own.

Inheriting Methods

An object has access not only to the methods defined for its class, but also to methods defined for its superclass, and for its superclass's superclass, all the way back to the root of the hierarchy. For instance, a Square object can use methods defined in the Rectangle, Shape, Graphic, and `NSObject` classes as well as methods defined in its own class.

Any new class you define in your program can therefore make use of the code written for all the classes above it in the hierarchy. This type of inheritance is a major benefit of object-oriented programming. When you use one of the object-oriented frameworks provided by Cocoa, your programs can take advantage of the basic functionality coded into the framework classes. You have to add only the code that customizes the standard functionality to your application.

Class objects also inherit from the classes above them in the hierarchy. But because they don't have instance variables (only instances do), they inherit only methods.

Overriding One Method With Another

There's one useful exception to inheritance: When you define a new class, you can implement a new method with the same name as one defined in a class farther up the hierarchy. The new method overrides the original; instances of the new class perform it rather than the original, and subclasses of the new class inherit it rather than the original.

For example, Graphic defines a `display` method that Rectangle overrides by defining its own version of `display`. The Graphic method is available to all kinds of objects that inherit from the Graphic class—but not to Rectangle objects, which instead perform the Rectangle version of `display`.

Although overriding a method blocks the original version from being inherited, other methods defined in the new class can skip over the redefined method and find the original (see "Messages to self and super" (page 43) to learn how).

A redefined method can also incorporate the very method it overrides. When it does, the new method serves only to refine or modify the method it overrides, rather than replace it outright. When several classes in the hierarchy define the same method, but each new version incorporates the version it overrides, the implementation of the method is effectively spread over all the classes.

Although a subclass can override inherited methods, it can't override inherited instance variables. Since an object has memory allocated for every instance variable it inherits, you can't override an inherited variable by declaring a new one with the same name. If you try, the compiler will complain.

Abstract Classes

Some classes are designed only or primarily so that other classes can inherit from them. These **abstract classes** group methods and instance variables that can be used by a number of different subclasses into a common definition. The abstract class is typically incomplete by itself, but contains useful code that reduces the implementation burden of its subclasses. (Because abstract classes must have subclasses to be useful, they're sometimes also called **abstract superclasses**.)

Unlike some other languages, Objective-C does not have syntax to mark classes as abstract, nor does it prevent you from creating an instance of an abstract class.

The `NSObject` class is the canonical example of an abstract class in Cocoa. You never use instances of the `NSObject` class in an application—it wouldn't be good for anything; it would be a generic object with the ability to do nothing in particular.

The `NSView` class, on the other hand, provides an example of an abstract class instances of which you might occasionally use directly.

Abstract classes often contain code that helps define the structure of an application. When you create subclasses of these classes, instances of your new classes fit effortlessly into the application structure and work automatically with other objects.

Class Types

A class definition is a specification for a kind of object. The class, in effect, defines a data type. The type is based not just on the data structure the class defines (instance variables), but also on the behavior included in the definition (methods).

A class name can appear in source code wherever a type specifier is permitted in C—for example, as an argument to the `sizeof` operator:

```
int i = sizeof(Rectangle);
```

Static Typing

You can use a class name in place of `id` to designate an object's type:

```
Rectangle *myRectangle;
```

Because this way of declaring an object type gives the compiler information about the kind of object it is, it's known as **static typing**. Just as `id` is actually a pointer, objects are statically typed as pointers to a class. Objects are always typed by a pointer. Static typing makes the pointer explicit; `id` hides it.

Static typing permits the compiler to do some type checking—for example, to warn if an object could receive a message that it appears not to be able to respond to—and to loosen some restrictions that apply to objects generically typed `id`. In addition, it can make your intentions clearer to others who read your source code. However, it doesn't defeat dynamic binding or alter the dynamic determination of a receiver's class at runtime.

An object can be statically typed to its own class or to any class that it inherits from. For example, since inheritance makes a Rectangle a kind of Graphic, a Rectangle instance could be statically typed to the Graphic class:

```
Graphic *myRectangle;
```

This is possible because a Rectangle is a Graphic. It's more than a Graphic since it also has the instance variables and method capabilities of a Shape and a Rectangle, but it's a Graphic nonetheless. For purposes of type checking, the compiler considers `myRectangle` to be a Graphic, but at runtime it's treated as a Rectangle.

See "Enabling Static Behavior" (page 91) for more on static typing and its benefits.

Type Introspection

Instances can reveal their types at runtime. The `isMemberOfClass:` method, defined in the `NSObject` class, checks whether the receiver is an instance of a particular class:

```
if ( [anObject isMemberOfClass:someClass] )
    ...
```

The `isKindOfClass:` method, also defined in the `NSObject` class, checks more generally whether the receiver inherits from or is a member of a particular class (whether it has the class in its inheritance path):

```
if ( [anObject isKindOfClass:someClass] )
    ...
```

The set of classes for which `isKindOfClass:` returns `YES` is the same set to which the receiver can be statically typed.

Introspection isn't limited to type information. Later sections of this chapter discuss methods that return the class object, report whether an object can respond to a message, and reveal other information.

See the `NSObject` class specification in the Foundation framework reference for more on `isKindOfClass:`, `isMemberOfClass:`, and related methods.

Class Objects

A class definition contains various kinds of information, much of it about instances of the class:

- The name of the class and its superclass
- A template describing a set of instance variables
- The declarations of method names and their return and argument types
- The method implementations

This information is compiled and recorded in data structures made available to the runtime system. The compiler creates just one object, a **class object**, to represent the class. The class object has access to all the information about the class, which means mainly information about what instances of the class are like. It's able to produce new instances according to the plan put forward in the class definition.

Although a class object keeps the prototype of a class instance, it's not an instance itself. It has no instance variables of its own and it can't perform methods intended for instances of the class. However, a class definition can include methods intended specifically for the class object—**class methods** as opposed to **instance methods**. A class object inherits class methods from the classes above it in the hierarchy, just as instances inherit instance methods.

In source code, the class object is represented by the class name. In the following example, the Rectangle class returns the class version number using a method inherited from the `NSObject` class:

```
int versionNumber = [Rectangle version];
```

However, the class name stands for the class object only as the receiver in a message expression. Elsewhere, you need to ask an instance or the class to return the class `id`. Both respond to a `class` message:

```
id aClass = [anObject class];
```

```
id rectClass = [Rectangle class];
```

As these examples show, class objects can, like all other objects, be typed `id`. But class objects can also be more specifically typed to the Class data type:

```
Class aClass = [anObject class];
Class rectClass = [Rectangle class];
```

All class objects are of type Class. Using this type name for a class is equivalent to using the class name to statically type an instance.

Class objects are thus full-fledged objects that can be dynamically typed, receive messages, and inherit methods from other classes. They're special only in that they're created by the compiler, lack data structures (instance variables) of their own other than those built from the class definition, and are the agents for producing instances at runtime.

Note: The compiler also builds a "metaclass object" for each class. It describes the class object just as the class object describes instances of the class. But while you can send messages to instances and to the class object, the metaclass object is used only internally by the runtime system.

Creating Instances

A principal function of a class object is to create new instances. This code tells the Rectangle class to create a new Rectangle instance and assign it to the `myRectangle` variable:

```
id  myRectangle;
myRectangle = [Rectangle alloc];
```

The `alloc` method dynamically allocates memory for the new object's instance variables and initializes them all to `0`—all, that is, except the `isa` variable that connects the new instance to its class. For an object to be useful, it generally needs to be more completely initialized. That's the function of an `init` method. Initialization typically follows immediately after allocation:

```
myRectangle = [[Rectangle alloc] init];
```

This line of code, or one like it, would be necessary before `myRectangle` could receive any of the messages that were illustrated in previous examples in this chapter. The `alloc` method returns a new instance and that instance performs an `init` method to set its initial state. Every class object has at least one method (like `alloc`) that enables it to produce new objects, and every instance has at least one method (like `init`) that prepares it for use. Initialization methods often take arguments to allow particular values to be passed and have keywords to label the arguments (`initWithPosition:size:`, for example, is a method that might initialize a new Rectangle instance), but they all begin with "init".

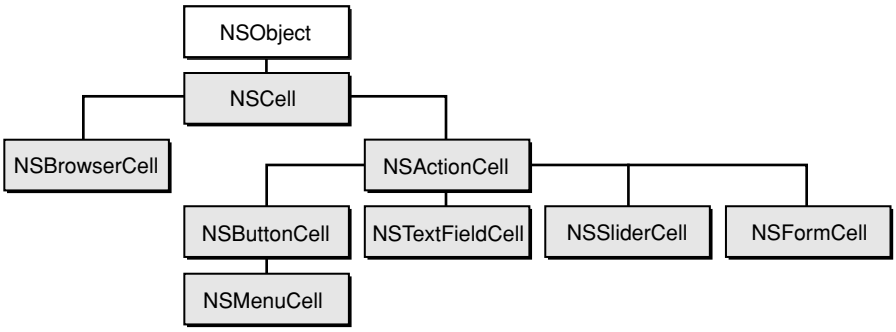
Customization With Class Objects

It's not just a whim of the Objective-C language that classes are treated as objects. It's a choice that has intended, and sometimes surprising, benefits for design. It's possible, for example, to customize an object with a class, where the class belongs to an open-ended set. In the Application Kit, for example, an `NSMatrix` object can be customized with a particular kind of `NSCell` object.

An NSMatrix object can take responsibility for creating the individual objects that represent its cells. It can do this when the matrix is first initialized and later when new cells are needed. The visible matrix that an NSMatrix object draws on the screen can grow and shrink at runtime, perhaps in response to user actions. When it grows, the matrix needs to be able to produce new objects to fill the new slots that are added.

But what kind of objects should they be? Each matrix displays just one kind of NSCell, but there are many different kinds. The inheritance hierarchy in Figure 1-3 shows some of those provided by the Application Kit. All inherit from the generic NSCell class:

Figure 1-3 Inheritance hierarchy for NSCell



When a matrix creates NSCell objects, should they be NSButtonCell objects to display a bank of buttons or switches, NSTextFieldCell objects to display fields where the user can enter and edit text, or some other kind of NSCell? The NSMatrix object must allow for any kind of cell, even types that haven't been invented yet.

One solution to this problem is to define the NSMatrix class as an abstract class and require everyone who uses it to declare a subclass and implement the methods that produce new cells. Because they would be implementing the methods, users of the class could be sure that the objects they created were of the right type.

But this requires others to do work that ought to be done in the NSMatrix class, and it unnecessarily proliferates the number of classes. Since an application might need more than one kind of NSMatrix, each with a different kind of NSCell, it could become cluttered with NSMatrix subclasses. Every time you invented a new kind of NSCell, you'd also have to define a new kind of NSMatrix. Moreover, programmers on different projects would be writing virtually identical code to do the same job, all to make up for NSMatrix's failure to do it.

A better solution, the solution the NSMatrix class actually adopts, is to allow NSMatrix instances to be initialized with a kind of NSCell—with a class object. It defines a setCellClass: method that passes the class object for the kind of NSCell object an NSMatrix should use to fill empty slots:

```
[myMatrix setCellClass:[NSButtonCell class]];
```

The NSMatrix object uses the class object to produce new cells when it's first initialized and whenever it's resized to contain more cells. This kind of customization would be difficult if classes weren't objects that could be passed in messages and assigned to variables.

Variables and Class Objects

When you define a new class, you can specify instance variables. Every instance of the class can maintain its own copy of the variables you declare—each object controls its own data. There is, however, no "class variable" counterpart to an instance variable. Only internal data structures, initialized from the class definition, are provided for the class. Moreover, a class object has no access to the instance variables of any instances; it can't initialize, read, or alter them.

For all the instances of a class to share data, you must define an external variable of some sort. The simplest way to do this is to declare a variable in the class implementation file as illustrated in the following code fragment.

```
int MCLSGlobalVariable;

@implementation MyClass
// implementation continues
```

In a more sophisticated implementation, you can declare a variable to be `static`, and provide class methods to manage it. Declaring a variable `static` limits its scope to just the class—and to just the part of the class that's implemented in the file. (Thus unlike instance variables, static variables cannot be inherited by, or directly manipulated by, subclasses.) This pattern is commonly used to define shared instances of a class (such as singletons, see "Creating a Singleton Instance" in *Cocoa Fundamentals Guide*).

```
static MyClass *MCLSSharedInstance;

@implementation MyClass

+ (MyClass *)sharedInstance
{
    // check for existence of shared instance
    // create if necessary
    return MCLSSharedInstance;
}
// implementation continues
```

Static variables help give the class object more functionality than just that of a "factory" producing instances; it can approach being a complete and versatile object in its own right. A class object can be used to coordinate the instances it creates, dispense instances from lists of objects already created, or manage other processes essential to the application. In the case when you need only one object of a particular class, you can put all the object's state into static variables and use only class methods. This saves the step of allocating and initializing an instance.

Note: It is also possible to use external variables that are not declared `static`, but the limited scope of static variables better serves the purpose of encapsulating data into separate objects.

Initializing a Class Object

If you want to use a class object for anything besides allocating instances, you may need to initialize it just as you would an instance. Although programs don't allocate class objects, Objective-C does provide a way for programs to initialize them.

If a class makes use of static or global variables, the `initialize` method is a good place to set their initial values. For example, if a class maintains an array of instances, the `initialize` method could set up the array and even allocate one or two default instances to have them ready.

The runtime system sends an `initialize` message to every class object before the class receives any other messages and after its superclass has received the `initialize` message. This gives the class a chance to set up its runtime environment before it's used. If no initialization is required, you don't need to write an `initialize` method to respond to the message.

Because of inheritance, an `initialize` message sent to a class that doesn't implement the `initialize` method is forwarded to the superclass, even though the superclass has already received the `initialize` message. For example, assume class A implements the `initialize` method, and class B inherits from class A but does not implement the `initialize` method. Just before class B is to receive its first message, the runtime system sends `initialize` to it. But, because class B doesn't implement `initialize`, class A's `initialize` is executed instead. Therefore, class A should ensure that its initialization logic is performed only once, and for the appropriate class.

To avoid performing initialization logic more than once, use the template in Listing 1-3 when implementing the `initialize` method.

Listing 1-3 Implementation of the initialize method

```
+ (void)initialize
{
  if (self == [ThisClass class]) {
        // Perform initialization here.
        ...
    }
}
```

Note: Remember that the runtime system sends `initialize` to each class individually. Therefore, in a class's implementation of the `initialize` method, you must not send the `initialize` message to its superclass.

Methods of the Root Class

All objects, classes and instances alike, need an interface to the runtime system. Both class objects and instances should be able to introspect about their abilities and to report their place in the inheritance hierarchy. It's the province of the `NSObject` class to provide this interface.

So that `NSObject`'s methods don't have to be implemented twice—once to provide a runtime interface for instances and again to duplicate that interface for class objects—class objects are given special dispensation to perform instance methods defined in the root class. When a class object receives a message that it can't respond to with a class method, the runtime system determines whether there's a root instance method that can respond. The only instance methods that a class object can perform are those defined in the root class, and only if there's no class method that can do the job.

For more on this peculiar ability of class objects to perform root instance methods, see the `NSObject` class specification in the Foundation framework reference.

Class Names in Source Code

In source code, class names can be used in only two very different contexts. These contexts reflect the dual role of a class as a data type and as an object:

- The class name can be used as a type name for a kind of object. For example:

  ```
  Rectangle *anObject;
  ```

  Here `anObject` is statically typed to be a pointer to a Rectangle. The compiler expects it to have the data structure of a Rectangle instance and the instance methods defined and inherited by the Rectangle class. Static typing enables the compiler to do better type checking and makes source code more self-documenting. See "Enabling Static Behavior" (page 91) for details.

  Only instances can be statically typed; class objects can't be, since they aren't members of a class, but rather belong to the Class data type.

- As the receiver in a message expression, the class name refers to the class object. This usage was illustrated in several of the earlier examples. The class name can stand for the class object only as a message receiver. In any other context, you must ask the class object to reveal its `id` (by sending it a `class` message). The example below passes the Rectangle class as an argument in an `isKindOfClass:` message.

  ```
  if ( [anObject isKindOfClass:[Rectangle class]] )
      ...
  ```

  It would have been illegal to simply use the name "Rectangle" as the argument. The class name can only be a receiver.

  If you don't know the class name at compile time but have it as a string at runtime, you can use `NSClassFromString` to return the class object:

  ```
  NSString *className;
      ...
  if ( [anObject isKindOfClass:NSClassFromString(className)] )
      ...
  ```

  This function returns `nil` if the string it's passed is not a valid class name.

Classnames exist in the same namespace as global variables and function names. A class and a global variable can't have the same name. Classnames are about the only names with global visibility in Objective-C.

Testing Class Equality

You can test two class objects for equality using a direct pointer comparison. It is important, though, to get the correct class. There are several features in the Cocoa frameworks that dynamically and transparently subclass existing classes to extend their functionality (for example, key-value observing and Core Data—see *Key-Value Observing Programming Guide* and *Core Data Programming Guide* respectively). When this happens, the `class` method is typically overridden such that the dynamic subclass masquerades as the class it replaces. When testing for class equality, you should therefore compare the values returned by the `class` method rather than those returned by lower-level functions. Put in terms of API:

```
[object class] != object_getClass(object) != *((Class*)object)
```

You should therefore test two classes for equality as follows:

```
if ([objectA class] == [objectB class]) { //...
```

Defining a Class

Much of object-oriented programming consists of writing the code for new objects—defining new classes. In Objective-C, classes are defined in two parts:

- An **interface** that declares the methods and instance variables of the class and names its superclass
- An **implementation** that actually defines the class (contains the code that implements its methods)

These are typically split between two files, sometimes however a class definition may span several files through the use of a feature called a "category." Categories can compartmentalize a class definition or extend an existing one. Categories are described in "Categories and Extensions" (page 79).

Source Files

Although the compiler doesn't require it, the interface and implementation are usually separated into two different files. The interface file must be made available to anyone who uses the class.

A single file can declare or implement more than one class. Nevertheless, it's customary to have a separate interface file for each class, if not also a separate implementation file. Keeping class interfaces separate better reflects their status as independent entities.

Interface and implementation files typically are named after the class. The name of the implementation file has the `.m` extension, indicating that it contains Objective-C source code. The interface file can be assigned any other extension. Because it's included in other source files, the name of the interface file usually has the `.h` extension typical of header files. For example, the Rectangle class would be declared in `Rectangle.h` and defined in `Rectangle.m`.

Separating an object's interface from its implementation fits well with the design of object-oriented programs. An object is a self-contained entity that can be viewed from the outside almost as a "black box." Once you've determined how an object interacts with other elements in your program—that is, once you've declared its interface—you can freely alter its implementation without affecting any other part of the application.

Class Interface

The declaration of a class interface begins with the compiler directive `@interface` and ends with the directive `@end`. (All Objective-C directives to the compiler begin with "@".)

```
@interface ClassName : ItsSuperclass
{
    instance variable declarations
}
method declarations
@end
```

The first line of the declaration presents the new class name and links it to its superclass. The superclass defines the position of the new class in the inheritance hierarchy, as discussed under "Inheritance" (page 24). If the colon and superclass name are omitted, the new class is declared as a root class, a rival to the `NSObject` class.

Following the first part of the class declaration, braces enclose declarations of **instance variables**, the data structures that are part of each instance of the class. Here's a partial list of instance variables that might be declared in the Rectangle class:

```
float width;
float height;
BOOL filled;
NSColor *fillColor;
```

Methods for the class are declared next, after the braces enclosing instance variables and before the end of the class declaration. The names of methods that can be used by class objects, **class methods**, are preceded by a plus sign:

```
+ alloc;
```

The methods that instances of a class can use, **instance methods**, are marked with a minus sign:

```
- (void)display;
```

Although it's not a common practice, you can define a class method and an instance method with the same name. A method can also have the same name as an instance variable. This is more common, especially if the method returns the value in the variable. For example, Circle has a `radius` method that could match a `radius` instance variable.

Method return types are declared using the standard C syntax for casting one type to another:

```
- (float)radius;
```

Argument types are declared in the same way:

```
- (void)setRadius:(float)aRadius;
```

If a return or argument type isn't explicitly declared, it's assumed to be the default type for methods and messages—an `id`. The `alloc` method illustrated earlier returns `id`.

When there's more than one argument, the arguments are declared within the method name after the colons. Arguments break the name apart in the declaration, just as in a message. For example:

```
- (void)setWidth:(float)width height:(float)height;
```

Methods that take a variable number of arguments declare them using a comma and ellipsis points, just as a function would:

```
- makeGroup:group, ...;
```

Importing the Interface

The interface file must be included in any source module that depends on the class interface—that includes any module that creates an instance of the class, sends a message to invoke a method declared for the class, or mentions an instance variable declared in the class. The interface is usually included with the `#import` directive:

```
#import "Rectangle.h"
```

This directive is identical to `#include`, except that it makes sure that the same file is never included more than once. It's therefore preferred and is used in place of `#include` in code examples throughout Objective-C–based documentation.

To reflect the fact that a class definition builds on the definitions of inherited classes, an interface file begins by importing the interface for its superclass:

```
#import "ItsSuperclass.h"

@interface ClassName : ItsSuperclass
{
    instance variable declarations
}
method declarations
@end
```

This convention means that every interface file includes, indirectly, the interface files for all inherited classes. When a source module imports a class interface, it gets interfaces for the entire inheritance hierarchy that the class is built upon.

Note that if there is a *precomp*—a precompiled header—that supports the superclass, you may prefer to import the precomp instead.

Referring to Other Classes

An interface file declares a class and, by importing its superclass, implicitly contains declarations for all inherited classes, from `NSObject` on down through its superclass. If the interface mentions classes not in this hierarchy, it must import them explicitly or declare them with the `@class` directive:

```
@class Rectangle, Circle;
```

This directive simply informs the compiler that "Rectangle" and "Circle" are class names. It doesn't import their interface files.

An interface file mentions class names when it statically types instance variables, return values, and arguments. For example, this declaration

```
- (void)setPrimaryColor:(NSColor *)aColor;
```

mentions the `NSColor` class.

Since declarations like this simply use the class name as a type and don't depend on any details of the class interface (its methods and instance variables), the `@class` directive gives the compiler sufficient forewarning of what to expect. However, where the interface to a class is actually used (instances created, messages sent),

the class interface must be imported. Typically, an interface file uses `@class` to declare classes, and the corresponding implementation file imports their interfaces (since it will need to create instances of those classes or send them messages).

The `@class` directive minimizes the amount of code seen by the compiler and linker, and is therefore the simplest way to give a forward declaration of a class name. Being simple, it avoids potential problems that may come with importing files that import still other files. For example, if one class declares a statically typed instance variable of another class, and their two interface files import each other, neither class may compile correctly.

The Role of the Interface

The purpose of the interface file is to declare the new class to other source modules (and to other programmers). It contains all the information they need to work with the class (programmers might also appreciate a little documentation).

- The interface file tells users how the class is connected into the inheritance hierarchy and what other classes—inherited or simply referred to somewhere in the class—are needed.
- The interface file also lets the compiler know what instance variables an object contains, and tells programmers what variables subclasses inherit. Although instance variables are most naturally viewed as a matter of the implementation of a class rather than its interface, they must nevertheless be declared in the interface file. This is because the compiler must be aware of the structure of an object where it's used, not just where it's defined. As a programmer, however, you can generally ignore the instance variables of the classes you use, except when defining a subclass.
- Finally, through its list of method declarations, the interface file lets other modules know what messages can be sent to the class object and instances of the class. Every method that can be used outside the class definition is declared in the interface file; methods that are internal to the class implementation can be omitted.

Class Implementation

The definition of a class is structured very much like its declaration. It begins with the `@implementation` directive and ends with the `@end` directive:

```
@implementation ClassName : ItsSuperclass
{
    instance variable declarations
}
method definitions
@end
```

However, every implementation file must import its own interface. For example, `Rectangle.m` imports `Rectangle.h`. Because the implementation doesn't need to repeat any of the declarations it imports, it can safely omit:

- The name of the superclass
- The declarations of instance variables

This simplifies the implementation and makes it mainly devoted to method definitions:

```
#import "ClassName.h"

@implementation ClassName
method definitions
@end
```

Methods for a class are defined, like C functions, within a pair of braces. Before the braces, they're declared in the same manner as in the interface file, but without the semicolon. For example:

```
+ (id)alloc
{
    ...
}

- (BOOL)isFilled
{
    ...
}

- (void)setFilled:(BOOL)flag
{
    ...
}
```

Methods that take a variable number of arguments handle them just as a function would:

```
#import <stdarg.h>

 ...

- getGroup:group, ...
{
    va_list ap;
    va_start(ap, group);
    ...
}
```

Referring to Instance Variables

By default, the definition of an instance method has all the instance variables of the object within its scope. It can refer to them simply by name. Although the compiler creates the equivalent of C structures to store instance variables, the exact nature of the structure is hidden. You don't need either of the structure operators (`.` or `->`) to refer to an object's data. For example, the following method definition refers to the receiver's `filled` instance variable:

```
- (void)setFilled:(BOOL)flag
{
    filled = flag;
    ...
}
```

Neither the receiving object nor its `filled` instance variable is declared as an argument to this method, yet the instance variable falls within its scope. This simplification of method syntax is a significant shorthand in the writing of Objective-C code.

When the instance variable belongs to an object that's not the receiver, the object's type must be made explicit to the compiler through static typing. In referring to the instance variable of a statically typed object, the structure pointer operator (`->`) is used.

Suppose, for example, that the Sibling class declares a statically typed object, `twin`, as an instance variable:

```
@interface Sibling : NSObject
{
    Sibling *twin;
    int gender;
    struct features *appearance;
}
```

As long as the instance variables of the statically typed object are within the scope of the class (as they are here because `twin` is typed to the same class), a Sibling method can set them directly:

```
- makeIdenticalTwin
{
    if ( !twin ) {
        twin = [[Sibling alloc] init];
        twin->gender = gender;
        twin->appearance = appearance;
    }
    return twin;
}
```

The Scope of Instance Variables

Although they're declared in the class interface, instance variables are more a matter of the way a class is implemented than of the way it's used. An object's interface lies in its methods, not in its internal data structures.

Often there's a one-to-one correspondence between a method and an instance variable, as in the following example:

```
- (BOOL)isFilled
{
    return filled;
}
```

But this need not be the case. Some methods might return information not stored in instance variables, and some instance variables might store information that an object is unwilling to reveal.

As a class is revised from time to time, the choice of instance variables may change, even though the methods it declares remain the same. As long as messages are the vehicle for interacting with instances of the class, these changes won't really affect its interface.

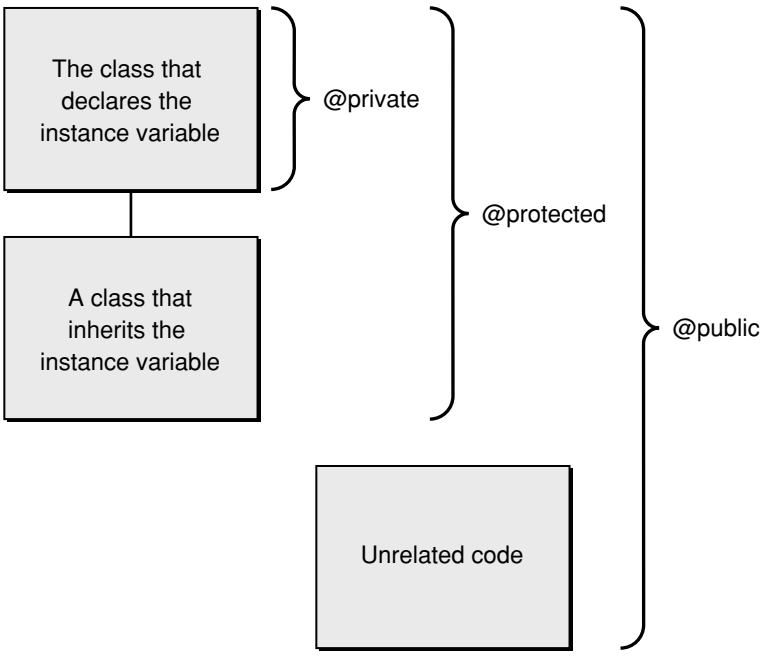
To enforce the ability of an object to hide its data, the compiler limits the scope of instance variables—that is, limits their visibility within the program. But to provide flexibility, it also lets you explicitly set the scope at three different levels. Each level is marked by a compiler directive:

Directive	Meaning
`@private`	The instance variable is accessible only within the class that declares it.

Directive	Meaning
`@protected`	The instance variable is accessible within the class that declares it and within classes that inherit it.
`@public`	The instance variable is accessible everywhere.
`@package`	Using the modern runtime, an `@package` instance variable acts like `@public` inside the image that implements the class, but `@private` outside. This is analogous to `private_extern` for variables and functions. Any code outside the class implementation's image that tries to use the instance variable will get a link error. This is most useful for instance variables in framework classes, where `@private` may be too restrictive but `@protected` or `@public` too permissive.

This is illustrated in Figure 2-1.

Figure 2-1 The scope of instance variables



A directive applies to all the instance variables listed after it, up to the next directive or the end of the list. In the following example, the `age` and `evaluation` instance variables are private, `name`, `job`, and `wage` are protected, and `boss` is public.

```
@interface Worker : NSObject
{
    char *name;
@private
    int age;
    char *evaluation;
@protected
    id job;
    float wage;
```

```
@public
    id boss;
}
```

By default, all unmarked instance variables (like `name` above) are `@protected`.

All instance variables that a class declares, no matter how they're marked, are within the scope of the class definition. For example, a class that declares a `job` instance variable, such as the Worker class shown above, can refer to it in a method definition:

```
- promoteTo:newPosition
{
    id old = job;
    job = newPosition;
    return old;
}
```

Obviously, if a class couldn't access its own instance variables, the instance variables would be of no use whatsoever.

Normally, a class also has access to the instance variables it inherits. The ability to refer to an instance variable is usually inherited along with the variable. It makes sense for classes to have their entire data structures within their scope, especially if you think of a class definition as merely an elaboration of the classes it inherits from. The `promoteTo:` method illustrated earlier could just as well have been defined in any class that inherits the `job` instance variable from the Worker class.

However, there are reasons why you might want to restrict inheriting classes from directly accessing an instance variable:

- Once a subclass accesses an inherited instance variable, the class that declares the variable is tied to that part of its implementation. In later versions, it can't eliminate the variable or alter the role it plays without inadvertently breaking the subclass.
- Moreover, if a subclass accesses an inherited instance variable and alters its value, it may inadvertently introduce bugs in the class that declares the variable, especially if the variable is involved in class-internal dependencies.

To limit an instance variable's scope to just the class that declares it, you must mark it `@private`. Instance variables marked `@private` are only available to subclasses by calling public accessor methods, if they exist.

At the other extreme, marking a variable `@public` makes it generally available, even outside of class definitions that inherit or declare the variable. Normally, to get information stored in an instance variable, other objects must send a message requesting it. However, a public instance variable can be accessed anywhere as if it were a field in a C structure. For example:

```
Worker *ceo = [[Worker alloc] init];
ceo->boss = nil;
```

Note that the object must be statically typed.

Marking instance variables `@public` defeats the ability of an object to hide its data. It runs counter to a fundamental principle of object-oriented programming—the encapsulation of data within objects where it's protected from view and inadvertent error. Public instance variables should therefore be avoided except in extraordinary cases.

Messages to self and super

Objective-C provides two terms that can be used within a method definition to refer to the object that performs the method—`self` and `super`.

Suppose, for example, that you define a `reposition` method that needs to change the coordinates of whatever object it acts on. It can invoke the `setOrigin::` method to make the change. All it needs to do is send a `setOrigin::` message to the same object that the `reposition` message itself was sent to. When you're writing the reposition code, you can refer to that object as either `self` or `super`. The `reposition` method could read either:

```
- reposition
{
    ...
    [self setOrigin:someX :someY];
    ...
}
```

or:

```
- reposition
{
    ...
    [super setOrigin:someX :someY];
    ...
}
```

Here, `self` and `super` both refer to the object receiving a `reposition` message, whatever object that may happen to be. The two terms are quite different, however. `self` is one of the hidden arguments that the messaging routine passes to every method; it's a local variable that can be used freely within a method implementation, just as the names of instance variables can be. `super` is a term that substitutes for `self` only as the receiver in a message expression. As receivers, the two terms differ principally in how they affect the messaging process:

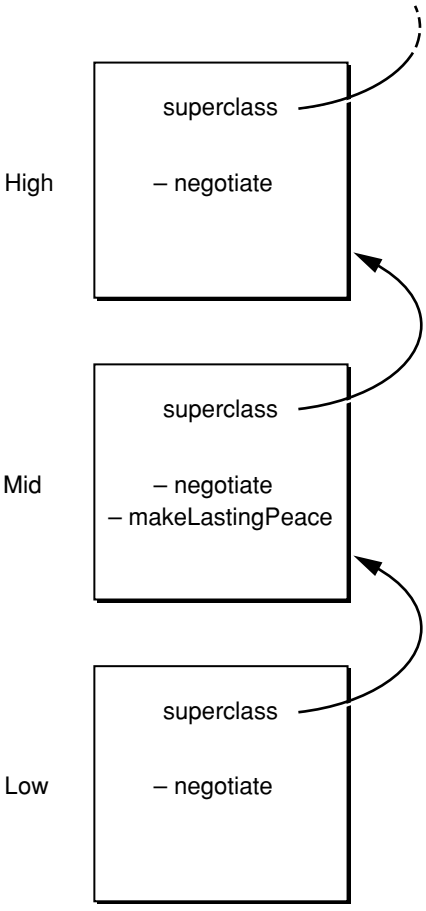
- `self` searches for the method implementation in the usual manner, starting in the dispatch table of the receiving object's class. In the example above, it would begin with the class of the object receiving the reposition message.
- `super` starts the search for the method implementation in a very different place. It begins in the superclass of the class that defines the method where `super` appears. In the example above, it would begin with the superclass of the class where reposition is defined.

Wherever `super` receives a message, the compiler substitutes another messaging routine for the `objc_msgSend` function. The substitute routine looks directly to the superclass of the defining class—that is, to the superclass of the class sending the message to `super`—rather than to the class of the object receiving the message.

An Example

The difference between `self` and `super` becomes clear in a hierarchy of three classes. Suppose, for example, that we create an object belonging to a class called Low. Low's superclass is Mid; Mid's superclass is High. All three classes define a method called `negotiate`, which they use for a variety of purposes. In addition, Mid defines an ambitious method called `makeLastingPeace`, which also has need of the `negotiate` method. This is illustrated in Figure 2-2:

Figure 2-2 High, Mid, Low



We now send a message to our Low object to perform the `makeLastingPeace` method, and `makeLastingPeace`, in turn, sends a `negotiate` message to the same Low object. If source code calls this object `self`,

```
- makeLastingPeace
{
    [self negotiate];
    ...
}
```

the messaging routine finds the version of `negotiate` defined in Low, `self`'s class. However, if Mid's source code calls this object `super`,

```
- makeLastingPeace
{
    [super negotiate];
    ...
}
```

the messaging routine will find the version of `negotiate` defined in High. It ignores the receiving object's class (Low) and skips to the superclass of Mid, since Mid is where `makeLastingPeace` is defined. Neither message finds Mid's version of `negotiate`.

As this example illustrates, `super` provides a way to bypass a method that overrides another method. Here it enabled `makeLastingPeace` to avoid the Mid version of `negotiate` that redefined the original High version.

Not being able to reach Mid's version of `negotiate` may seem like a flaw, but, under the circumstances, it's right to avoid it:

- The author of the Low class intentionally overrode Mid's version of `negotiate` so that instances of the Low class (and its subclasses) would invoke the redefined version of the method instead. The designer of Low didn't want Low objects to perform the inherited method.
- In sending the message to `super`, the author of Mid's `makeLastingPeace` method intentionally skipped over Mid's version of `negotiate` (and over any versions that might be defined in classes like Low that inherit from Mid) to perform the version defined in the High class. Mid's designer wanted to use the High version of `negotiate` and no other.

Mid's version of `negotiate` could still be used, but it would take a direct message to a Mid instance to do it.

Using super

Messages to `super` allow method implementations to be distributed over more than one class. You can override an existing method to modify or add to it, and still incorporate the original method in the modification:

```
- negotiate
{
    ...
    return [super negotiate];
}
```

For some tasks, each class in the inheritance hierarchy can implement a method that does part of the job and passes the message on to `super` for the rest. The `init` method, which initializes a newly allocated instance, is designed to work like this. Each `init` method has responsibility for initializing the instance variables defined in its class. But before doing so, it sends an `init` message to `super` to have the classes it inherits from initialize their instance variables. Each version of `init` follows this procedure, so classes initialize their instance variables in the order of inheritance:

```
- (id)init
{
    if (self = [super init]) {
        ...
    }
}
```

Initializer methods have some additional constraints, and are described in more detail in "Allocating and Initializing Objects" (page 47).

It's also possible to concentrate core functionality in one method defined in a superclass, and have subclasses incorporate the method through messages to `super`. For example, every class method that creates an instance must allocate storage for the new object and initialize its `isa` variable to the class structure. This is typically left to the `alloc` and `allocWithZone:` methods defined in the `NSObject` class. If another class overrides these methods (a rare case), it can still get the basic functionality by sending a message to `super`.

Redefining self

`super` is simply a flag to the compiler telling it where to begin searching for the method to perform; it's used only as the receiver of a message. But `self` is a variable name that can be used in any number of ways, even assigned a new value.

There's a tendency to do just that in definitions of class methods. Class methods are often concerned not with the class object, but with instances of the class. For example, many class methods combine allocation and initialization of an instance, often setting up instance variable values at the same time. In such a method, it might be tempting to send messages to the newly allocated instance and to call the instance `self`, just as in an instance method. But that would be an error. `self` and `super` both refer to the receiving object—the object that gets a message telling it to perform the method. Inside an instance method, `self` refers to the instance; but inside a class method, `self` refers to the class object. This is an example of what not to do:

```
+ (Rectangle *)rectangleOfColor:(NSColor *) color
{
    self = [[Rectangle alloc] init]; // BAD
    [self setColor:color];
    return [self autorelease];
}
```

To avoid confusion, it's usually better to use a variable other than `self` to refer to an instance inside a class method:

```
+ (id)rectangleOfColor:(NSColor *)color
{
    id newInstance = [[Rectangle alloc] init]; // GOOD
    [newInstance setColor:color];
    return [newInstance autorelease];
}
```

In fact, rather than sending the `alloc` message to the class in a class method, it's often better to send `alloc` to `self`. This way, if the class is subclassed, and the `rectangleOfColor:` message is received by a subclass, the instance returned will be the same type as the subclass (for example, the `array` method of `NSArray` is inherited by `NSMutableArray`).

```
+ (id)rectangleOfColor:(NSColor *)color
{
    id newInstance = [[self alloc] init]; // EXCELLENT
    [newInstance setColor:color];
    return [newInstance autorelease];
}
```

See "Allocating and Initializing Objects" (page 47) for more information about object allocation.

Allocating and Initializing Objects

Allocating and Initializing Objects

It takes two steps to create an object using Objective-C. You must:

- Dynamically allocate memory for the new object
- Initialize the newly allocated memory to appropriate values

An object isn't fully functional until both steps have been completed. Each step is accomplished by a separate method but typically in a single line of code:

```
id anObject = [[Rectangle alloc] init];
```

Separating allocation from initialization gives you individual control over each step so that each can be modified independently of the other. The following sections look first at allocation and then at initialization, and discuss how they are controlled and modified.

In Objective-C, memory for new objects is allocated using class methods defined in the `NSObject` class. `NSObject` defines two principal methods for this purpose, `alloc` and `allocWithZone:`.

These methods allocate enough memory to hold all the instance variables for an object belonging to the receiving class. They don't need to be overridden and modified in subclasses.

The `alloc` and `allocWithZone:` methods initialize a newly allocated object's `isa` instance variable so that it points to the object's class (the class object). All other instance variables are set to `0`. Usually, an object needs to be more specifically initialized before it can be safely used.

This initialization is the responsibility of class-specific instance methods that, by convention, begin with the abbreviation "init". If the method takes no arguments, the method name is just those four letters, `init`. If it takes arguments, labels for the arguments follow the "init" prefix. For example, an `NSView` object can be initialized with an `initWithFrame:` method.

Every class that declares instance variables must provide an `init...` method to initialize them. The `NSObject` class declares the `isa` variable and defines an `init` method. However, since `isa` is initialized when memory for an object is allocated, all `NSObject`'s `init` method does is return `self`. `NSObject` declares the method mainly to establish the naming convention described earlier.

The Returned Object

An `init...` method normally initializes the instance variables of the receiver, then returns it. It's the responsibility of the method to return an object that can be used without error.

However, in some cases, this responsibility can mean returning a different object than the receiver. For example, if a class keeps a list of named objects, it might provide an `initWithName:` method to initialize new instances. If there can be no more than one object per name, `initWithName:` might refuse to assign the same name to two objects. When asked to assign a new instance a name that's already being used by another object, it might free the newly allocated instance and return the other object—thus ensuring the uniqueness of the name while at the same time providing what was asked for, an instance with the requested name.

In a few cases, it might be impossible for an `init...` method to do what it's asked to do. For example, an `initFromFile:` method might get the data it needs from a file passed as an argument. If the file name it's passed doesn't correspond to an actual file, it won't be able to complete the initialization. In such a case, the `init...` method could free the receiver and return `nil`, indicating that the requested object can't be created.

Because an `init...` method might return an object other than the newly allocated receiver, or even return `nil`, it's important that programs use the value returned by the initialization method, not just that returned by `alloc` or `allocWithZone:`. The following code is very dangerous, since it ignores the return of `init`.

```
id anObject = [SomeClass alloc];
[anObject init];
[anObject someOtherMessage];
```

Instead, to safely initialize an object, you should combine allocation and initialization messages in one line of code.

```
id anObject = [[SomeClass alloc] init];
[anObject someOtherMessage];
```

If there's a chance that the `init...` method might return `nil` (see "Handling Initialization Failure" (page 50)), then you should check the return value before proceeding:

```
id anObject = [[SomeClass alloc] init];
if ( anObject )
    [anObject someOtherMessage];
else
    ...
```

Implementing an Initializer

When a new object is created, all bits of memory (except for `isa`)—and hence the values for all its instance variables—are set to `0`. In some situations, this may be all you require when an object is initialized; in many others, you want to provide other default values for an object's instance variables, or you want to pass values as arguments to the initializer. In these other cases, you need to write a custom initializer. In Objective-C, custom initializers are subject to more constraints and conventions than are most other methods.

Constraints and Conventions

There are several constraints and conventions that apply to initializer methods that do not apply to other methods:

- By convention, the name of a custom initializer method begins with `init`.

Examples from the Foundation framework include, `initWithFormat:`, `initWithObjects:`, and `initWithObjectsAndKeys:`.

- The return type of an initializer method should be `id`.

  The reason for this is that `id` gives an indication that the class is purposefully not considered—that the class is unspecified and subject to change, depending on context of invocation. For example, `NSString` provides a method `initWithFormat:`. When sent to an instance of `NSMutableString` (a subclass of `NSString`), however, the message returns an instance of `NSMutableString`, not `NSString`. (See also, though, the singleton example given in "Combining Allocation and Initialization" (page 55).)

- In the implementation of a custom initializer, you must ultimately invoke a **designated initializer**.

  Designated initializers are described in "The Designated Initializer" (page 53); a full explanation of this issue is given in "Coordinating Classes" (page 51).

  In brief, if you are implementing a new designated initializer, it must invoke the superclass' designated initializer. If you are implementing any other initializer, it should invoke its own class's designated initializer, or another of its own initializers that ultimately invokes the designated initializer.

  By default (such as with `NSObject`), the designated initializer is `init`.

- You should assign `self` to the value returned by the initializer.

  This is because the initializer could return a different object than the original receiver.

- If you set the value of an instance variable, you typically do so using direct assignment rather than using an accessor method.

  This avoids the possibility of triggering unwanted side-effects in the accessors.

- At the end of the initializer, you must return `self`, unless the initializer fails in which case you return `nil`.

  Failed initializers are discussed in more detail in "Handling Initialization Failure" (page 50).

The following example illustrates the implementation of a custom initializer for a class that inherits from `NSObject` and has an instance variable, `creationDate`, that represents the time when the object was created:

```
- (id)init {
    // Assign self to value returned by super's designated initializer
    // Designated initializer for NSObject is init
    if (self = [super init]) {
        creationDate = [[NSDate alloc] init];
    }
    return self;
}
```

(The reason for using the `if (self = [super init])` pattern is discussed in "Handling Initialization Failure" (page 50).)

An initializer doesn't need to provide an argument for each variable. For example, if a class requires its instances to have a name and a data source, it might provide an `initWithName:fromURL:` method, but set nonessential instance variables to arbitrary values or allow them to have the null values set by default. It could then rely on methods like `setEnabled:`, `setFriend:`, and `setDimensions:` to modify default values after the initialization phase had been completed.

The next example illustrates the implementation of a custom initializer that takes a single argument. In this case, the class inherits from `NSView`. It shows that you can do work before invoking the super class's designated initializer.

```
- (id)initWithImage:(NSImage *)anImage {

    // Find the size for the new instance from the image
    NSSize size = anImage.size;
    NSRect frame = NSMakeRect(0.0, 0.0, size.width, size.height);

    // Assign self to value returned by super's designated initializer
    // Designated initializer for NSView is initWithFrame:
    if (self = [super initWithFrame:frame]) {

        image = [anImage retain];
    }
    return self;
}
```

This example doesn't show what to do if there are any problems during initialization; this is discussed in the next section.

Handling Initialization Failure

In general, if there is a problem during an initialization method, you should call `[self release]` and return `nil`.

There are two main consequences of this policy:

- Any object (whether your own class, a subclass, or an external caller) that receives a `nil` from an initializer method should be able to deal with it. In the unlikely case where the caller has established any external references to the object before the call, this includes undoing any connections.
- You must make sure that `dealloc` methods are safe in presence of partially-initialized objects.

Note: You should only call `[self release]` at the point of failure. If you get `nil` back from an invocation of the superclass's initializer, you should not also call `release`. You should simply clean up any references you set up that are not dealt with in `dealloc` and return `nil`. This is typically handled by the pattern of performing initialization within a block dependent on a test of the return value of the superclass's initializer—as seen in previous examples:

```
- (id)init {
    if (self = [super init]) {
        creationDate = [[NSDate alloc] init];
    }
    return self;
}
```

The following example builds on that shown in "Constraints and Conventions" (page 48) to show how to handle an inappropriate value passed as the parameter:

```
- (id)initWithImage:(NSImage *)anImage {
```

```
    if (anImage == nil) {
        [self release];
        return nil;
    }

    // Find the size for the new instance from the image
    NSSize size = anImage.size;
    NSRect frame = NSMakeRect(0.0, 0.0, size.width, size.height);

    // Assign self to value returned by super's designated initializer
    // Designated initializer for NSView is initWithFrame:
    if (self = [super initWithFrame:frame]) {

        image = [anImage retain];
    }
    return self;
}
```

The next example illustrates best practice where, in the case of a problem, there is a possibility of returning meaningful information in the form of an `NSError` object returned by reference:

```
- (id)initWithURL:(NSURL *)aURL error:(NSError **)errorPtr {

    if (self = [super init]) {

        NSData *data = [[NSData alloc] initWithContentsOfURL:aURL
                                       options:NSUncachedRead error:errorPtr];

        if (data == nil) {
           // In this case the error object is created in the NSData initializer
            [self release];
            return nil;
        }
        // implementation continues...
```

You should typically not use exceptions to signify errors of this sort—for more information, see *Error Handling Programming Guide*.

Coordinating Classes

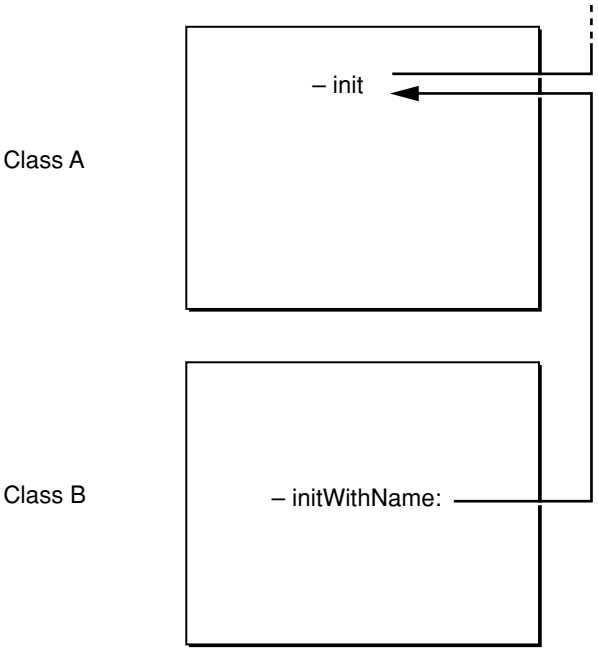
The `init...` methods a class defines typically initialize only those variables declared in that class. Inherited instance variables are initialized by sending a message to `super` to perform an initialization method defined somewhere farther up the inheritance hierarchy:

```
- (id)initWithName:(NSString *)string {
    if ( self = [super init] ) {
        name = [string copy];
    }
    return self;
}
```

The message to `super` chains together initialization methods in all inherited classes. Because it comes first, it ensures that superclass variables are initialized before those declared in subclasses. For example, a Rectangle object must be initialized as an `NSObject`, a Graphic, and a Shape before it's initialized as a Rectangle.

The connection between the `initWithName:` method illustrated above and the inherited `init` method it incorporates is illustrated in Figure 3-1:

Figure 3-1 Incorporating an Inherited Initialization Method

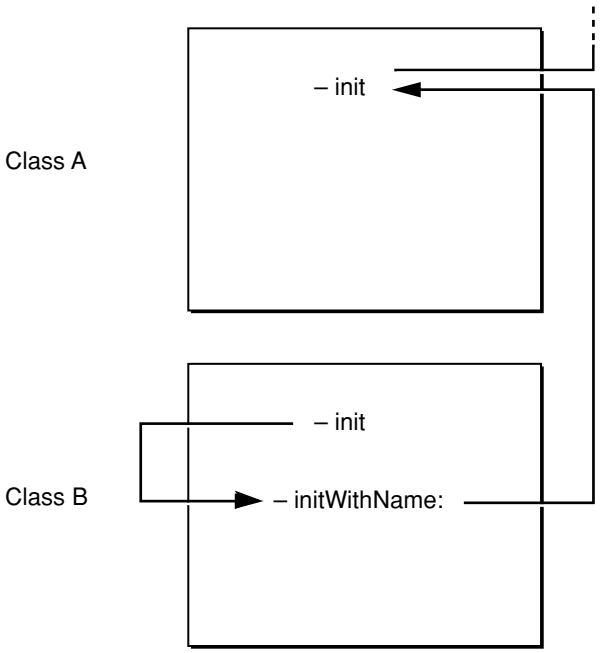


A class must also make sure that all inherited initialization methods work. For example, if class A defines an `init` method and its subclass B defines an `initWithName:` method, as shown in Figure 3-1, B must also make sure that an `init` message successfully initializes B instances. The easiest way to do that is to replace the inherited `init` method with a version that invokes `initWithName:`:

```
- init {
    return [self initWithName:"default"];
}
```

The `initWithName:` method would, in turn, invoke the inherited method, as shown earlier. Figure 3-2 includes B's version of `init`:

Figure 3-2 Covering an Inherited Initialization Model



Covering inherited initialization methods makes the class you define more portable to other applications. If you leave an inherited method uncovered, someone else may use it to produce incorrectly initialized instances of your class.

The Designated Initializer

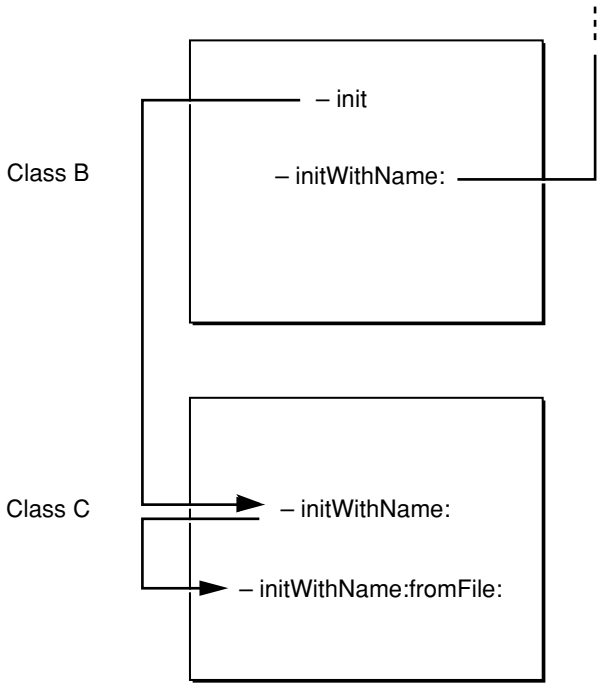
In the example given in "Coordinating Classes" (page 51), `initWithName:` would be the **designated initializer** for its class (class B). The designated initializer is the method in each class that guarantees inherited instance variables are initialized (by sending a message to `super` to perform an inherited method). It's also the method that does most of the work, and the one that other initialization methods in the same class invoke. It's a Cocoa convention that the designated initializer is always the method that allows the most freedom to determine the character of a new instance (usually this is the one with the most arguments, but not always).

It's important to know the designated initializer when defining a subclass. For example, suppose we define class C, a subclass of B, and implement an `initWithName:fromFile:` method. In addition to this method, we have to make sure that the inherited `init` and `initWithName:` methods also work for instances of C. This can be done just by covering B's `initWithName:` with a version that invokes `initWithName:fromFile:`.

```
- initWithName:(char *)string {
    return [self initWithName:string fromFile:NULL];
}
```

For an instance of the C class, the inherited `init` method invokes this new version of `initWithName:` which invokes `initWithName:fromFile:`. The relationship between these methods is shown in Figure 3-3:

Figure 3-3 Covering the Designated Initializer



This figure omits an important detail. The `initWithName:fromFile:` method, being the designated initializer for the C class, sends a message to `super` to invoke an inherited initialization method. But which of B's methods should it invoke, `init` or `initWithName:`? It can't invoke `init`, for two reasons:

- Circularity would result (`init` invokes C's `initWithName:`, which invokes `initWithName:fromFile:`, which invokes `init` again).
- It won't be able to take advantage of the initialization code in B's version of `initWithName:`.

Therefore, `initWithName:fromFile:` must invoke `initWithName:`:

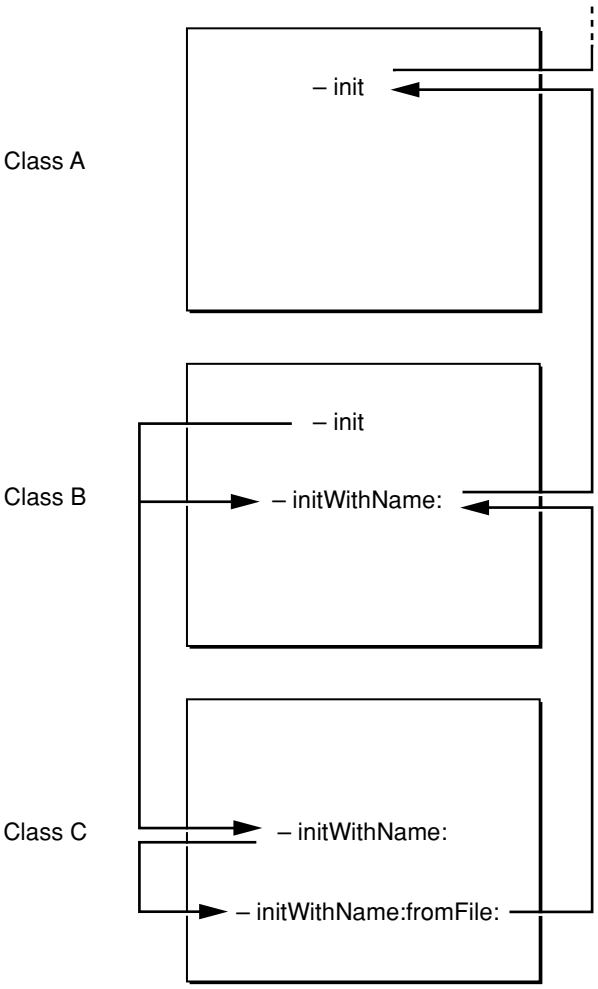
```
- initWithName:(char *)string fromFile:(char *)pathname {
    if ( self = [super initWithName:string] )
        ...
}
```

General Principle: The designated initializer in a class must, through a message to `super`, invoke the designated initializer in a superclass.

Designated initializers are chained to each other through messages to `super`, while other initialization methods are chained to designated initializers through messages to `self`.

Figure 3-4 shows how all the initialization methods in classes A, B, and C are linked. Messages to `self` are shown on the left and messages to `super` are shown on the right.

Figure 3-4 Initialization Chain



Note that B's version of `init` sends a message to `self` to invoke the `initWithName:` method. Therefore, when the receiver is an instance of the B class, it invokes B's version of `initWithName:`, and when the receiver is an instance of the C class, it invokes C's version.

Combining Allocation and Initialization

In Cocoa, some classes define creation methods that combine the two steps of allocating and initializing to return new, initialized instances of the class. These methods are often referred to as **convenience constructors** and typically take the form `+` *className...* where *className* is the name of the class. For example, `NSString` has the following methods (among others):

```
+ (id)stringWithCString:(const char *)cString encoding:(NSStringEncoding)enc;
+ (id)stringWithFormat:(NSString *)format, ...;
```

Similarly, `NSArray` defines the following class methods that combine allocation and initialization:

```
+ (id)array;
```

```
+ (id)arrayWithObject:(id)anObject;
+ (id)arrayWithObjects:(id)firstObj, ...;
```

Important: It is important to understand the memory management implications of using these methods if you do not use garbage collection (see "Memory Management" (page 15)). You must read *Memory Management Programming Guide* to understand the policy that applies to these convenience constructors.

Notice that the return type of these methods is `id`. This is for the same reason as for initializer methods, as discussed in "Constraints and Conventions" (page 48).

Methods that combine allocation and initialization are particularly valuable if the allocation must somehow be informed by the initialization. For example, if the data for the initialization is taken from a file, and the file might contain enough data to initialize more than one object, it would be impossible to know how many objects to allocate until the file is opened. In this case, you might implement a `listFromFile:` method that takes the name of the file as an argument. It would open the file, see how many objects to allocate, and create a List object large enough to hold all the new objects. It would then allocate and initialize the objects from data in the file, put them in the List, and finally return the List.

It also makes sense to combine allocation and initialization in a single method if you want to avoid the step of blindly allocating memory for a new object that you might not use. As mentioned in "The Returned Object" (page 47), an `init...` method might sometimes substitute another object for the receiver. For example, when `initWithName:` is passed a name that's already taken, it might free the receiver and in its place return the object that was previously assigned the name. This means, of course, that an object is allocated and freed immediately without ever being used.

If the code that determines whether the receiver should be initialized is placed inside the method that does the allocation instead of inside `init...`, you can avoid the step of allocating a new instance when one isn't needed.

In the following example, the `soloist` method ensures that there's no more than one instance of the Soloist class. It allocates and initializes a single shared instance:

```
+ (Soloist *)soloist {
    static Soloist *instance = nil;

    if ( instance == nil ) {
        instance = [[self alloc] init];
    }
    return instance;
}
```

Notice that in this case the return type is `Soloist *`. Since this method returns a singleton share instance, strong typing is appropriate—there is no expectation that this method will be overridden.

Protocols

Protocols declare methods that can be implemented by any class. Protocols are useful in at least three situations:

- To declare methods that others are expected to implement
- To declare the interface to an object while concealing its class
- To capture similarities among classes that are not hierarchically related

Declaring Interfaces for Others to Implement

Class and category interfaces declare methods that are associated with a particular class—mainly methods that the class implements. Informal and formal **protocols**, on the other hand, declare methods that are independent of any specific class, but which any class, and perhaps many classes, might implement.

A protocol is simply a list of method declarations, unattached to a class definition. For example, these methods that report user actions on the mouse could be gathered into a protocol:

```
- (void)mouseDown:(NSEvent *)theEvent;
- (void)mouseDragged:(NSEvent *)theEvent;
- (void)mouseUp:(NSEvent *)theEvent;
```

Any class that wanted to respond to mouse events could adopt the protocol and implement its methods.

Protocols free method declarations from dependency on the class hierarchy, so they can be used in ways that classes and categories cannot. Protocols list methods that are (or may be) implemented somewhere, but the identity of the class that implements them is not of interest. What is of interest is whether or not a particular class **conforms** to the protocol—whether it has implementations of the methods the protocol declares. Thus objects can be grouped into types not just on the basis of similarities due to the fact that they inherit from the same class, but also on the basis of their similarity in conforming to the same protocol. Classes in unrelated branches of the inheritance hierarchy might be typed alike because they conform to the same protocol.

Protocols can play a significant role in object-oriented design, especially where a project is divided among many implementors or it incorporates objects developed in other projects. Cocoa software uses protocols heavily to support interprocess communication through Objective-C messages.

However, an Objective-C program doesn't need to use protocols. Unlike class definitions and message expressions, they're optional. Some Cocoa frameworks use them; some don't. It all depends on the task at hand.

Methods for Others to Implement

If you know the class of an object, you can look at its interface declaration (and the interface declarations of the classes it inherits from) to find what messages it responds to. These declarations advertise the messages it can receive. Protocols provide a way for it to also advertise the messages it sends.

Communication works both ways; objects send messages as well as receive them. For example, an object might delegate responsibility for a certain operation to another object, or it may on occasion simply need to ask another object for information. In some cases, an object might be willing to notify other objects of its actions so that they can take whatever collateral measures might be required.

If you develop the class of the sender and the class of the receiver as part of the same project (or if someone else has supplied you with the receiver and its interface file), this communication is easily coordinated. The sender simply imports the interface file of the receiver. The imported file declares the method selectors the sender uses in the messages it sends.

However, if you develop an object that sends messages to objects that aren't yet defined—objects that you're leaving for others to implement—you won't have the receiver's interface file. You need another way to declare the methods you use in messages but don't implement. A protocol serves this purpose. It informs the compiler about methods the class uses and also informs other implementors of the methods they need to define to have their objects work with yours.

Suppose, for example, that you develop an object that asks for the assistance of another object by sending it `helpOut:` and other messages. You provide an `assistant` instance variable to record the outlet for these messages and define a companion method to set the instance variable. This method lets other objects register themselves as potential recipients of your object's messages:

```
- setAssistant:anObject
{
    assistant = anObject;
}
```

Then, whenever a message is to be sent to the `assistant`, a check is made to be sure that the receiver implements a method that can respond:

```
- (BOOL)doWork
{
    ...
    if ( [assistant respondsToSelector:@selector(helpOut:)] ) {
        [assistant helpOut:self];
        return YES;
    }
    return NO;
}
```

Since, at the time you write this code, you can't know what kind of object might register itself as the `assistant`, you can only declare a protocol for the `helpOut:` method; you can't import the interface file of the class that implements it.

Declaring Interfaces for Anonymous Objects

A protocol can be used to declare the methods of an **anonymous object**, an object of unknown class. An anonymous object may represent a service or handle a limited set of functions, especially where only one object of its kind is needed. (Objects that play a fundamental role in defining an application's architecture and objects that you must initialize before using are not good candidates for anonymity.)

Objects are not anonymous to their developers, of course, but they are anonymous when the developer supplies them to someone else. For example, consider the following situations:

- Someone who supplies a framework or a suite of objects for others to use can include objects that are not identified by a class name or an interface file. Lacking the name and class interface, users have no way of creating instances of the class. Instead, the supplier must provide a ready-made instance. Typically, a method in another class returns a usable object:

  ```
  id formatter = [receiver formattingService];
  ```

  The object returned by the method is an object without a class identity, at least not one the supplier is willing to reveal. For it to be of any use at all, the supplier must be willing to identify at least some of the messages that it can respond to. This is done by associating the object with a list of methods declared in a protocol.

- You can send Objective-C messages to **remote objects**—objects in other applications. (Remote Messaging (page 105) in the *Objective-C Runtime Programming Guide*, discusses this possibility in more detail.)

  Each application has its own structure, classes, and internal logic. But you don't need to know how another application works or what its components are to communicate with it. As an outsider, all you need to know is what messages you can send (the protocol) and where to send them (the receiver).

  An application that publishes one of its objects as a potential receiver of remote messages must also publish a protocol declaring the methods the object will use to respond to those messages. It doesn't have to disclose anything else about the object. The sending application doesn't need to know the class of the object or use the class in its own design. All it needs is the protocol.

Protocols make anonymous objects possible. Without a protocol, there would be no way to declare an interface to an object without identifying its class.

Note: Even though the supplier of an anonymous object doesn't reveal its class, the object itself reveals it at runtime. A class message returns the anonymous object's class. However, there's usually little point in discovering this extra information; the information in the protocol is sufficient.

Non-Hierarchical Similarities

If more than one class implements a set of methods, those classes are often grouped under an abstract class that declares the methods they have in common. Each subclass may re-implement the methods in its own way, but the inheritance hierarchy and the common declaration in the abstract class captures the essential similarity between the subclasses.

However, sometimes it's not possible to group common methods in an abstract class. Classes that are unrelated in most respects might nevertheless need to implement some similar methods. This limited similarity may not justify a hierarchical relationship. For example, you might want to add support for creating XML representations of objects in your application and for initializing objects from an XML representation:

```
- (NSXMLElement *)XMLRepresentation;
- initFromXMLRepresentation:(NSXMLElement *)xmlString;
```

These methods could be grouped into a protocol and the similarity between implementing classes accounted for by noting that they all conform to the same protocol.

Objects can be typed by this similarity (the protocols they conform to), rather than by their class. For example, an `NSMatrix` instance must communicate with the objects that represent its cells. The matrix could require each of these objects to be a kind of `NSCell` (a type based on class) and rely on the fact that all objects that inherit from the `NSCell` class have the methods needed to respond to `NSMatrix` messages. Alternatively, the `NSMatrix` object could require objects representing cells to have methods that can respond to a particular set of messages (a type based on protocol). In this case, the `NSMatrix` object wouldn't care what class a cell object belonged to, just that it implemented the methods.

Formal Protocols

The Objective-C language provides a way to formally declare a list of methods (including declared properties) as a protocol. Formal protocols are supported by the language and the runtime system. For example, the compiler can check for types based on protocols, and objects can introspect at runtime to report whether or not they conform to a protocol.

Declaring a Protocol

You declare formal protocols with the `@protocol` directive:

```
@protocol ProtocolName
method declarations
@end
```

For example, you could declare an XML representation protocol like this:

```
@protocol MyXMLSupport
- initFromXMLRepresentation:(NSXMLElement *)XMLElement;
- (NSXMLElement *)XMLRepresentation;
@end
```

Unlike class names, protocol names don't have global visibility. They live in their own namespace.

Optional Protocol Methods

Protocol methods can be marked as optional using the `@optional` keyword. Corresponding to the `@optional` modal keyword, there is a `@required` keyword to formally denote the semantics of the default behavior. You can use `@optional` and `@required` to partition your protocol into sections as you see fit. If you do not specify any keyword, the default is `@required`.

```
@protocol MyProtocol

- (void)requiredMethod;

@optional
- (void)anOptionalMethod;
- (void)anotherOptionalMethod;

@required
- (void)anotherRequiredMethod;

@end
```

Note: In Mac OS X v10.5, protocols may not include optional declared properties. This constraint is removed in Mac OS X v10.6 and later.

Informal Protocols

In addition to formal protocols, you can also define an **informal** protocol by grouping the methods in a category declaration:

```
@interface NSObject ( MyXMLSupport )
- initFromXMLRepresentation:(NSXMLElement *)XMLElement;
- (NSXMLElement *)XMLRepresentation;
@end
```

Informal protocols are typically declared as categories of the `NSObject` class, since that broadly associates the method names with any class that inherits from `NSObject`. Because all classes inherit from the root class, the methods aren't restricted to any part of the inheritance hierarchy. (It would also be possible to declare an informal protocol as a category of another class to limit it to a certain branch of the inheritance hierarchy, but there is little reason to do so.)

When used to declare a protocol, a category interface doesn't have a corresponding implementation. Instead, classes that implement the protocol declare the methods again in their own interface files and define them along with other methods in their implementation files.

An informal protocol bends the rules of category declarations to list a group of methods but not associate them with any particular class or implementation.

Being informal, protocols declared in categories don't receive much language support. There's no type checking at compile time nor a check at runtime to see whether an object conforms to the protocol. To get these benefits, you must use a formal protocol. An informal protocol may be useful when all the methods are optional, such as for a delegate, but (on Mac OS X v10.5 and later) it is typically better to use a formal protocol with optional methods.

Protocol Objects

Just as classes are represented at runtime by class objects and methods by selector codes, formal protocols are represented by a special data type—instances of the Protocol class. Source code that deals with a protocol (other than to use it in a type specification) must refer to the Protocol object.

In many ways, protocols are similar to class definitions. They both declare methods, and at runtime they're both represented by objects—classes by class objects and protocols by Protocol objects. Like class objects, Protocol objects are created automatically from the definitions and declarations found in source code and are used by the runtime system. They're not allocated and initialized in program source code.

Source code can refer to a Protocol object using the `@protocol()` directive—the same directive that declares a protocol, except that here it has a set of trailing parentheses. The parentheses enclose the protocol name:

```
Protocol *myXMLSupportProtocol = @protocol(MyXMLSupport);
```

This is the only way that source code can conjure up a Protocol object. Unlike a class name, a protocol name doesn't designate the object—except inside `@protocol()`.

The compiler creates a Protocol object for each protocol declaration it encounters, but only if the protocol is also:

- Adopted by a class, or
- Referred to somewhere in source code (using `@protocol()`)

Protocols that are declared but not used (except for type checking as described below) aren't represented by Protocol objects at runtime.

Adopting a Protocol

Adopting a protocol is similar in some ways to declaring a superclass. Both assign methods to the class. The superclass declaration assigns it inherited methods; the protocol assigns it methods declared in the protocol list. A class is said to **adopt** a formal protocol if in its declaration it lists the protocol within angle brackets after the superclass name:

```
@interface ClassName : ItsSuperclass < protocol list >
```

Categories adopt protocols in much the same way:

```
@interface ClassName ( CategoryName ) < protocol list >
```

A class can adopt more than one protocol; names in the protocol list are separated by commas.

```
@interface Formatter : NSObject < Formatting, Prettifying >
```

A class or category that adopts a protocol must implement all the required methods the protocol declares, otherwise the compiler issues a warning. The Formatter class above would define all the required methods declared in the two protocols it adopts, in addition to any it might have declared itself.

A class or category that adopts a protocol must import the header file where the protocol is declared. The methods declared in the adopted protocol are not declared elsewhere in the class or category interface.

It's possible for a class to simply adopt protocols and declare no other methods. For example, the following class declaration adopts the Formatting and Prettifying protocols, but declares no instance variables or methods of its own:

```
@interface Formatter : NSObject < Formatting, Prettifying >
@end
```

Conforming to a Protocol

A class is said to **conform** to a formal protocol if it adopts the protocol or inherits from another class that adopts it. An instance of a class is said to conform to the same set of protocols its class conforms to.

Since a class must implement all the required methods declared in the protocols it adopts, saying that a class or an instance conforms to a protocol is equivalent to saying that it has in its repertoire all the methods the protocol declares.

It's possible to check whether an object conforms to a protocol by sending it a `conformsToProtocol:` message.

```
if ( ! [receiver conformsToProtocol:@protocol(MyXMLSupport)]  ) {
    // Object does not conform to MyXMLSupport protocol
    // If you are expecting receiver to implement methods declared in the
    //  MyXMLSupport protocol, this is probably an error
}
```

(Note that there is also a class method with the same name—`conformsToProtocol:`.)

The `conformsToProtocol:` test is like the `respondsToSelector:` test for a single method, except that it tests whether a protocol has been adopted (and presumably all the methods it declares implemented) rather than just whether one particular method has been implemented. Because it checks for all the methods in the protocol, `conformsToProtocol:` can be more efficient than `respondsToSelector:`.

The `conformsToProtocol:` test is also like the `isKindOfClass:` test, except that it tests for a type based on a protocol rather than a type based on the inheritance hierarchy.

Type Checking

Type declarations for objects can be extended to include formal protocols. Protocols thus offer the possibility of another level of type checking by the compiler, one that's more abstract since it's not tied to particular implementations.

In a type declaration, protocol names are listed between angle brackets after the type name:

```
- (id <Formatting>)formattingService;
id <MyXMLSupport> anObject;
```

Just as static typing permits the compiler to test for a type based on the class hierarchy, this syntax permits the compiler to test for a type based on conformance to a protocol.

For example, if Formatter is an abstract class, this declaration

```
Formatter *anObject;
```

groups all objects that inherit from Formatter into a type and permits the compiler to check assignments against that type.

Similarly, this declaration,

```
id <Formatting> anObject;
```

groups all objects that conform to the Formatting protocol into a type, regardless of their positions in the class hierarchy. The compiler can make sure only objects that conform to the protocol are assigned to the type.

In each case, the type groups similar objects—either because they share a common inheritance, or because they converge on a common set of methods.

The two types can be combined in a single declaration:

```
Formatter <Formatting> *anObject;
```

Protocols can't be used to type class objects. Only instances can be statically typed to a protocol, just as only instances can be statically typed to a class. (However, at runtime, both classes and instances will respond to a `conformsToProtocol:` message.)

Protocols Within Protocols

One protocol can incorporate other protocols using the same syntax that classes use to adopt a protocol:

```
@protocol ProtocolName < protocol list >
```

All the protocols listed between angle brackets are considered part of the *ProtocolName* protocol. For example, if the Paging protocol incorporates the Formatting protocol,

```
@protocol Paging < Formatting >
```

any object that conforms to the Paging protocol also conforms to Formatting. Type declarations

```
id <Paging> someObject;
```

and `conformsToProtocol:` messages

```
if ( [anotherObject conformsToProtocol:@protocol(Paging)] )
    ...
```

need to mention only the Paging protocol to test for conformance to Formatting as well.

When a class adopts a protocol, it must implement the required methods the protocol declares, as mentioned earlier. In addition, it must conform to any protocols the adopted protocol incorporates. If an incorporated protocol incorporates still other protocols, the class must also conform to them. A class can conform to an incorporated protocol by either:

- Implementing the methods the protocol declares, or
- Inheriting from a class that adopts the protocol and implements the methods.

Suppose, for example, that the Pager class adopts the Paging protocol. If Pager is a subclass of `NSObject`,

```
@interface Pager : NSObject < Paging >
```

it must implement all the Paging methods, including those declared in the incorporated Formatting protocol. It adopts the Formatting protocol along with Paging.

On the other hand, if Pager is a subclass of Formatter (a class that independently adopts the Formatting protocol),

```
@interface Pager : Formatter < Paging >
```

it must implement all the methods declared in the Paging protocol proper, but not those declared in Formatting. Pager inherits conformance to the Formatting protocol from Formatter.

Note that a class can conform to a protocol without formally adopting it simply by implementing the methods declared in the protocol.

Referring to Other Protocols

When working on complex applications, you occasionally find yourself writing code that looks like this:

```
#import "B.h"

@protocol A
- foo:(id <B>)anObject;
@end
```

where protocol B is declared like this:

```
#import "A.h"

@protocol B
- bar:(id <A>)anObject;
@end
```

In such a situation, circularity results and neither file will compile correctly. To break this recursive cycle, you must use the `@protocol` directive to make a forward reference to the needed protocol instead of importing the interface file where the protocol is defined. The following code excerpt illustrates how you would do this:

```
@protocol B;

@protocol A
- foo:(id <B>)anObject;
@end
```

Note that using the `@protocol` directive in this manner simply informs the compiler that “B” is a protocol to be defined later. It doesn’t import the interface file where protocol B is defined.

Declared Properties

The Objective-C "declared properties" feature provides a simple way to declare and implement an object's accessor methods.

Overview

There are two aspects to this language feature: the syntactic elements you use to specify and optionally synthesize **declared properties**, and a related syntactic element that is described in "Dot Syntax" (page 20).

You typically access an object's properties (in the sense of its attributes and relationships) through a pair of accessor (getter/setter) methods. By using accessor methods, you adhere to the principle of encapsulation (see "Mechanisms Of Abstraction" in *Object-Oriented Programming with Objective-C* > The Object Model). You can exercise tight control of the behavior of the getter/setter pair and the underlying state management while clients of the API remain insulated from the implementation changes.

Although using accessor methods has significant advantages, writing accessor methods is nevertheless a tedious process—particularly if you have to write code to support both garbage collected and reference counted environments. Moreover, aspects of the property that may be important to consumers of the API are left obscured—such as whether the accessor methods are thread-safe or whether new values are copied when set.

Declared properties address the problems with standard accessor methods by providing the following features:

- The property declaration provides a clear, explicit specification of how the accessor methods behave.
- The compiler can synthesize accessor methods for you, according to the specification you provide in the declaration. This means you have less code to write and maintain.
- Properties are represented syntactically as identifiers and are scoped, so the compiler can detect use of undeclared properties.

Property Declaration and Implementation

There are two parts to a declared property, its declaration and its implementation.

Property Declaration

A property declaration begins with the keyword `@property`. `@property` can appear anywhere in the method declaration list found in the `@interface` of a class. `@property` can also appear in the declaration of a protocol or category.

`@property(attributes) type name;`

`@property` declares a property. An optional parenthesized set of attributes provides additional details about the storage semantics and other behaviors of the property—see "Property Declaration Attributes" (page 68) for possible values. Like any other Objective-C type, each property has a type specification and a name.

Listing 5-1 illustrates the declaration of a simple property.

Listing 5-1 Declaring a simple property

```
@interface MyClass : NSObject
{
    float value;
}
@property float value;
@end
```

You can think of a property declaration as being equivalent to declaring two accessor methods. Thus

```
@property float value;
```

is equivalent to:

```
- (float)value;
- (void)setValue:(float)newValue;
```

A property declaration, however, provides additional information about how the accessor methods are implemented (as described in "Property Declaration Attributes" (page 68)).

Property Declaration Attributes

You can decorate a property with attributes by using the form `@property(attribute [, attribute2, ...])`. Like methods, properties are scoped to their enclosing interface declaration. For property declarations that use a comma delimited list of variable names, the property attributes apply to all of the named properties.

If you use the `@synthesize` directive to tell the compiler to create the accessor method(s), the code it generates matches the specification given by the keywords. If you implement the accessor method(s) yourself, you should ensure that it matches the specification (for example, if you specify `copy` you must make sure that you do copy the input value in the setter method).

Accessor Method Names

The default names for the getter and setter methods associated with a property are *propertyName* and `set`*PropertyName*`:` respectively—for example, given a property "foo", the accessors would be `foo` and `setFoo:`. The following attributes allow you to specify custom names instead. They are both optional and may appear with any other attribute (except for `readonly` in the case of `setter=`).

`getter=getterName`

Specifies the name of the get accessor for the property. The getter must return a type matching the property's type and take no arguments.

`setter=setterName`

Specifies the name of the set accessor for the property. The setter method must take a single argument of a type matching the property's type and must return `void`.

If you specify that a property is `readonly` then also specify a setter with `setter=`, you will get a compiler warning.

Typically you should specify accessor method names that are key-value coding compliant (see *Key-Value Coding Programming Guide*)—a common reason for using the `getter` decorator is to adhere to the `is`*PropertyName* convention for Boolean values.

Writability

These attributes specify whether or not a property has an associated set accessor. They are mutually exclusive.

`readwrite`

Indicates that the property should be treated as read/write. This is the default.

Both a getter and setter method will be required in the `@implementation`. If you use `@synthesize` in the implementation block, the getter and setter methods are synthesized.

`readonly`

Indicates that the property is read-only.

If you specify `readonly`, only a getter method is required in the `@implementation`. If you use `@synthesize` in the implementation block, only the getter method is synthesized. Moreover, if you attempt to assign a value using the dot syntax, you get a compiler error.

Setter Semantics

These attributes specify the semantics of a set accessor. They are mutually exclusive.

`assign`

Specifies that the setter uses simple assignment. This is the default.

You typically use this attribute for scalar types such as `NSInteger` and `CGRect`, or (in a reference-counted environment) for objects you don't own such as delegates.

`retain` and `assign` are effectively the same in a garbage-collected environment.

`retain`

Specifies that `retain` should be invoked on the object upon assignment. (The default is `assign`.)

The previous value is sent a `release` message.

Prior to Mac OS X v10.6, this attribute is valid only for Objective-C object types (so you cannot specify `retain` for Core Foundation objects—see "Core Foundation" (page 74)).

On Mac OS X v10.6 and later, you can use the __attribute__ keyword to specify that a Core Foundation property should be treated like an Objective-C object for memory management, as illustrated in this example:

```
@property(retain) __attribute__((NSObject)) CFDictionaryRef myDictionary;
```

`copy`

Specifies that a copy of the object should be used for assignment. (The default is `assign`.)

The previous value is sent a `release` message.

The copy is made by invoking the `copy` method. This attribute is valid only for object types, which must implement the `NSCopying` protocol. For further discussion, see "Copy" (page 73).

Different constraints apply depending on whether or not you use garbage collection:

- If you do not use garbage collection, for object properties you must explicitly specify one of `assign`, `retain` or `copy`—otherwise you will get a compiler warning. (This encourages you to think about what memory management behavior you want and type it explicitly.)

  To decide which you should choose, you need to understand Cocoa's memory management policy (see *Memory Management Programming Guide*).
- If you use garbage collection, you don't get a warning if you use the default (that is, if you don't specify any of `assign`, `retain` or `copy`) unless the property's type is a class that conforms to `NSCopying`. The default is usually what you want; if the property type can be copied, however, to preserve encapsulation you often want to make a private copy of the object.

Atomicity

This attribute specifies that accessor methods are not atomic. (There is no keyword to denote atomic.)

`nonatomic`
: Specifies that accessors are non-atomic. *By default, accessors are atomic.*

Properties are atomic by default so that synthesized accessors provide robust access to properties in a multi-threaded environment—that is, the value returned from the getter or set via the setter is always fully retrieved or set regardless of what other threads are executing concurrently. For more details, see "Performance and Threading" (page 77).

If you do not specify `nonatomic`, then in a reference counted environment a synthesized get accessor for an object property uses a lock and retains and autoreleases the returned value—the implementation will be similar to the following:

```
[_internal lock]; // lock using an object-level lock
id result = [[value retain] autorelease];
[_internal unlock];
return result;
```

If you specify `nonatomic`, then a synthesized accessor for an object property simply returns the value directly.

Markup and Deprecation

Properties support the full range of C style decorators. Properties can be deprecated and support `__attribute__` style markup, as illustrated in the following example:

```
@property CGFloat x
AVAILABLE_MAC_OS_X_VERSION_10_1_AND_LATER_BUT_DEPRECATED_IN_MAC_OS_X_VERSION_10_4;
@property CGFloat y __attribute__((...));
```

If you want to specify that a property is an Interface Builder outlet, you can use the `IBOutlet` identifier:

```
@property (nonatomic, retain) IBOutlet NSButton *myButton;
```

`IBOutlet` is not, though, a formal part of the list of attributes.

If you use garbage collection, you can use the storage modifiers `__weak` and `__strong` in a property's declaration:

```
@property (nonatomic, retain) __weak Link *parent;
```

but again they are not a formal part of the list of attributes.

Property Implementation Directives

You can use the `@synthesize` and `@dynamic` directives in `@implementation` blocks to trigger specific compiler actions. Note that neither is *required* for any given `@property` declaration.

Important: If you do not specify either `@synthesize` or `@dynamic` for a particular property, you must provide a getter and setter (or just a getter in the case of a `readonly` property) method implementation for that property.

`@synthesize`

You use the `@synthesize` keyword to tell the compiler that it should synthesize the setter and/or getter methods for the property if you do not supply them within the `@implementation` block.

Listing 5-2 Using @synthesize

```
@interface MyClass : NSObject
{
    NSString *value;
}
@property(copy, readwrite) NSString *value;
@end

@implementation MyClass
@synthesize value;
@end
```

You can use the form `property=ivar` to indicate that a particular instance variable should be used for the property, for example:

```
@synthesize firstName, lastName, age = yearsOld;
```

This specifies that the accessor methods for `firstName`, `lastName`, and `age` should be synthesized and that the property `age` is represented by the instance variable `yearsOld`. Other aspects of the synthesized methods are determined by the optional attributes (see "Property Declaration Attributes" (page 68)).

Whether or not you specify the name of the instance variable, `@synthesize` can only use an instance variable from the current class, not a superclass.

There are differences in the behavior that depend on the runtime (see also "Runtime Difference" (page 78)):

- For the legacy runtimes, instance variables must already be declared in the `@interface` block of the current class. If an instance variable of the same name and compatible type as the property exists, it is used—otherwise, you get a compiler error.
- For the modern runtimes (see Runtime Versions and Platforms in *Objective-C Runtime Programming Guide*), instance variables are synthesized as needed. If an instance variable of the same name already exists, it is used.

`@dynamic`

You use the `@dynamic` keyword to tell the compiler that you will fulfill the API contract implied by a property either by providing method implementations directly or at runtime using other mechanisms such as dynamic loading of code or dynamic method resolution. The example shown in Listing 5-3 illustrates using direct method implementations—it is equivalent to the example given in Listing 5-2 (page 71).

Listing 5-3 Using @dynamic with direct method implementations

```
@interface MyClass : NSObject
{
    NSString *value;
}
@property(copy, readwrite) NSString *value;
@end

// assume using garbage collection
@implementation MyClass
@dynamic value;

- (NSString *)value {
    return value;
}

- (void)setValue:(NSString *)newValue {
    if (newValue != value) {
        value = [newValue copy];
    }
}
@end
```

Using Properties

Supported Types

You can declare a property for any Objective-C class, Core Foundation data type, or "plain old data" (POD) type (see C++ Language Note: POD Types). For constraints on using Core Foundation types, however, see "Core Foundation" (page 74).

Property Re-declaration

You can re-declare a property in a subclass, but (with the exception of `readonly` vs. `readwrite`) you must repeat its attributes in whole in the subclasses. The same holds true for a property declared in a category or protocol—while the property may be redeclared in a category or protocol, the property's attributes must be repeated in whole.

If you declare a property in one class as `readonly`, you can redeclare it as `readwrite` in a class extension (see "Extensions" (page 81)), a protocol, or a subclass—see "Subclassing with Properties" (page 76). In the case of a class extension redeclaration, the fact that the property was redeclared prior to any `@synthesize` statement will cause the setter to be synthesized. The ability to redeclare a read-only property as read/write

enables two common implementation patterns: a mutable subclass of an immutable class (`NSString`, `NSArray`, and `NSDictionary` are all examples) and a property that has public API that is `readonly` but a private `readwrite` implementation internal to the class. The following example shows using a class extension to provide a property that is declared as read-only in the public header but is redeclared privately as read/write.

```
// public header file
@interface MyObject : NSObject {
    NSString *language;
}
@property (readonly, copy) NSString *language;
@end

// private implementation file
@interface MyObject ()
@property (readwrite, copy) NSString *language;
@end

@implementation MyObject
@synthesize language;
@end
```

Copy

If you use the `copy` declaration attribute, you specify that a value is copied during assignment. If you synthesize the corresponding accessor, the synthesized method uses the `copy` method. This is useful for attributes such as string objects where there is a possibility that the new value passed in a setter may be mutable (for example, an instance of `NSMutableString`) and you want to ensure that your object has its own private immutable copy. For example, if you declare a property as follows:

```
@property (nonatomic, copy) NSString *string;
```

then the synthesized setter method is similar to the following:

```
-(void)setString:(NSString *)newString {
    if (string != newString) {
        [string release];
        string = [newString copy];
    }
}
```

Although this works well for strings, it may present a problem if the attribute is a collection such as an array or a set. Typically you want such collections to be mutable, but the `copy` method returns an *immutable* version of the collection. In this situation, you have to provide your own implementation of the setter method, as illustrated in the following example.

```
@interface MyClass : NSObject {
    NSMutableArray *myArray;
}
@property (nonatomic, copy) NSMutableArray *myArray;
@end

@implementation MyClass

@synthesize myArray;
```

```
- (void)setMyArray:(NSMutableArray *)newArray {
    if (myArray != newArray) {
        [myArray release];
        myArray = [newArray mutableCopy];
    }
}

@end
```

dealloc

Declared properties fundamentally take the place of accessor method declarations; when you synthesize a property, the compiler only creates any absent accessor methods. There is no direct interaction with the `dealloc` method—properties are *not* automatically released for you. Declared properties do, however, provide a useful way to cross-check the implementation of your `dealloc` method: you can look for all the property declarations in your header file and make sure that object properties not marked `assign` are released, and those marked `assign` are not released.

Note: Typically in a `dealloc` method you should release object instance variables directly (rather than invoking a set accessor and passing `nil` as the parameter), as illustrated in this example:

```
- (void)dealloc {
    [property release];
    [super dealloc];
}
```

If you are using the modern runtime and synthesizing the instance variable, however, you cannot access the instance variable directly, so you must invoke the accessor method:

```
- (void)dealloc {
    [self setProperty:nil];
    [super dealloc];
}
```

Core Foundation

As noted in "Property Declaration Attributes" (page 68), prior to Mac OS X v10.6 you cannot specify the `retain` attribute for non-object types. If, therefore, you declare a property whose type is a CFType and synthesize the accessors as illustrated in the following example:

```
@interface MyClass : NSObject
{
    CGImageRef myImage;
}
@property(readwrite) CGImageRef myImage;
@end

@implementation MyClass
@synthesize myImage;
@end
```

then in a reference counted environment the generated set accessor will simply assign the new value to the instance variable (the new value is not retained and the old value is not released). This is typically incorrect, so you should not synthesize the methods, you should implement them yourself.

In a garbage collected environment, if the variable is declared `__strong`:

```
...
__strong CGImageRef myImage;
...
@property CGImageRef myImage;
```

then the accessors are synthesized appropriately—the image will not be CFRetain'd, but the setter will trigger a write barrier.

Example

The following example illustrates the use of properties in several different ways:

- The Link protocol declares a property, `next`.
- MyClass adopts the Link protocol so implicitly also declares the property `next`. MyClass also declares several other properties.
- `creationTimestamp` and `next` are synthesized but use existing instance variables with different names;
- `name` is synthesized, and uses instance variable synthesis (recall that instance variable synthesis is not supported using the legacy runtime—see "Property Implementation Directives" (page 71) and "Runtime Difference" (page 78));
- `gratuitousFloat` has a `dynamic` directive—it is supported using direct method implementations;
- `nameAndAge` does not have a `dynamic` directive, but this is the default value; it is supported using a direct method implementation (since it is read-only, it only requires a getter) with a specified name (`nameAndAgeAsString`).

Listing 5-4 Declaring properties for a class

```
@protocol Link
@property id <Link> next;
@end


@interface MyClass : NSObject <Link>
{
    NSTimeInterval intervalSinceReferenceDate;
    CGFloat gratuitousFloat;
    id <Link> nextLink;
}
@property(readonly) NSTimeInterval creationTimestamp;
@property(copy) NSString *name;
@property CGFloat gratuitousFloat;
@property(readonly, getter=nameAndAgeAsString) NSString *nameAndAge;

@end
```

```
@implementation MyClass

@synthesize creationTimestamp = intervalSinceReferenceDate, name;
// Synthesizing 'name' is an error in legacy runtimes;
// in modern runtimes, the instance variable is synthesized.

@synthesize next = nextLink;
// Uses instance variable "nextLink" for storage.

@dynamic gratuitousFloat;
// This directive is not strictly necessary.

- (CGFloat)gratuitousFloat {
    return gratuitousFloat;
}
- (void)setGratuitousFloat:(CGFloat)aValue {
    gratuitousFloat = aValue;
}


- (NSString *)nameAndAgeAsString {
    return [NSString stringWithFormat:@"%@ (%fs)", [self name],
                [NSDate timeIntervalSinceReferenceDate] -
intervalSinceReferenceDate];
}


 - (id)init {
    if (self = [super init]) {
        intervalSinceReferenceDate = [NSDate timeIntervalSinceReferenceDate];
    }
    return self;
}

- (void)dealloc {
    [nextLink release];
    [name release];
    [super dealloc];
}

@end
```

Subclassing with Properties

You can override a `readonly` property to make it writable. For example, you could define a class `MyInteger` with a `readonly` property, `value`:

```
@interface MyInteger : NSObject
{
    NSInteger value;
}
@property(readonly) NSInteger value;
@end

@implementation MyInteger
```

```
@synthesize value;
@end
```

You could then implement a subclass, `MyMutableInteger`, which redefines the property to make it writable:

```
@interface MyMutableInteger : MyInteger
@property(readwrite) NSInteger value;
@end

@implementation MyMutableInteger
@dynamic value;

- (void)setValue:(NSInteger)newX {
    value = newX;
}
@end
```

Performance and Threading

If you supply your own method implementation, the fact that you declared a property has no effect on its efficiency or thread safety.

If you use synthesized properties, the method implementations generated by the compiler depend on the specification you supply. The declaration attributes that affect performance and threading are `retain`, `assign`, `copy`, and `nonatomic`. The first three of these affect only the implementation of the assignment part of the set method, as *illustrated* below (the implementation may not be exactly as shown):

```
// assign
property = newValue;

// retain
if (property != newValue) {
    [property release];
    property = [newValue retain];
}

// copy
if (property != newValue) {
    [property release];
    property = [newValue copy];
}
```

The effect of the `nonatomic` attribute depends on the environment. By default, the synthesized accessors are atomic. In a reference counted environment, guaranteeing atomic behavior requires the use of a lock; moreover a returned object is retained and autoreleased, as illustrated in "Atomicity" (page 70). If such accessors are invoked frequently, this may have a significant impact on performance. In a garbage collected environment, most synthesized methods are atomic without incurring this overhead.

It is important to understand that the goal of the atomic implementation is to provide *robust* accessors—it does not guarantee *correctness* of your code. Although "atomic" means that access to the *property* is thread-safe, simply making all the properties in your class atomic does not mean that your *class* or more generally your object graph is "thread safe"—thread safety cannot be expressed at the level of individual accessor methods. For more about multi-threading, see *Threading Programming Guide*.

Runtime Difference

In general the behavior of properties is identical on all runtimes (see Runtime Versions and Platforms in *Objective-C Runtime Programming Guide*). There is one key difference: the modern runtime supports instance variable synthesis whereas the legacy runtime does not.

For `@synthesize` to work in the legacy runtime, you must either provide an instance variable with the same name and compatible type of the property or specify another existing instance variable in the `@synthesize` statement. With the modern runtime, if you do not provide an instance variable, the compiler adds one for you. For example, given the following class declaration and implementation:

```
@interface MyClass : NSObject {
    float sameName;
    float otherName;
}
@property float sameName;
@property float differentName;
@property float noDeclaredIvar;
@end

@implementation MyClass
@synthesize sameName;
@synthesize differentName=otherName;
@synthesize noDeclaredIvar;
@end
```

the compiler for the legacy runtime would generate an error at `@synthesize noDeclaredIvar;` whereas the compiler for the modern runtime would add an instance variable to represent `noDeclaredIvar`.

Categories and Extensions

A category allows you to add methods to an existing class—even to one to which you do not have the source. This is a powerful feature that allows you to extend the functionality of existing classes without subclassing. Using categories, you can also split the implementation of your own classes between several files. Class extensions are similar, but allow additional *required* API to be declared for a class in locations other than within the primary class `@interface` block

Adding Methods to Classes

You can add methods to a class by declaring them in an interface file under a category name and defining them in an implementation file under the same name. The category name indicates that the methods are additions to a class declared elsewhere, not a new class. You cannot, however, use a category to add additional instance variables to a class.

The methods the category adds become part of the class type. For example, methods added to the `NSArray` class in a category are among the methods the compiler expects an `NSArray` instance to have in its repertoire. Methods added to the `NSArray` class in a subclass are not included in the `NSArray` type. (This matters only for statically typed objects, since static typing is the only way the compiler can know an object's class.)

Category methods can do anything that methods defined in the class proper can do. At runtime, there's no difference. The methods the category adds to the class are inherited by all the class's subclasses, just like other methods.

The declaration of a category interface looks very much like a class interface declaration—except the category name is listed within parentheses after the class name and the superclass isn't mentioned. Unless its methods don't access any instance variables of the class, the category must import the interface file for the class it extends:

```
#import "ClassName.h"

@interface ClassName ( CategoryName )
// method declarations
@end
```

The implementation, as usual, imports its own interface. A common naming convention is that the base file name of the category is the name of the class the category extends followed by "+" followed by the name of the category. A category implementation (in a file named `ClassName+CategoryName.m`) might therefore look like this:

```
#import "ClassName+CategoryName.h"

@implementation ClassName ( CategoryName )
// method definitions
@end
```

Note that a category can't declare additional instance variables for the class; it includes only methods. However, all instance variables within the scope of the class are also within the scope of the category. That includes all instance variables declared by the class, even ones declared `@private`.

There's no limit to the number of categories that you can add to a class, but each category name must be different, and each should declare and define a different set of methods.

How you Use Categories

There are several ways in which you can use categories:

- To extend classes defined by other implementors.

  For example, you can add methods to the classes defined in the Cocoa frameworks. The added methods are inherited by subclasses and are indistinguishable at runtime from the original methods of the class.

- As an alternative to a subclass.

  Rather than define a subclass to extend an existing class, through a category you can add methods to the class directly. For example, you could add categories to `NSArray` and other Cocoa classes. As in the case of a subclass, you don't need source code for the class you're extending.

- To distribute the implementation of a new class into separate source files.

  For example, you could group the methods of a large class into several categories and put each category in a different file. When used like this, categories can benefit the development process in a number of ways—they:

  - Provide a simple way of grouping related methods. Similar methods defined in different classes can be kept together in the same source file.
  - Simplify the management of a large class when several developers contribute to the class definition.
  - Let you achieve some of the benefits of incremental compilation for a very large class.
  - Can help improve locality of reference for commonly used methods.
  - Enable you to configure a class differently for separate applications, without having to maintain different versions of the same source code.

- To declare informal protocols.

  See "Informal Protocols " (page 61), as discussed under "Declaring Interfaces for Others to Implement" (page 57).

Although the language currently allows you to use a category to override methods the class inherits, or even methods declared in the class interface, you are strongly discouraged from using this functionality. A category is not a substitute for a subclass. There are several significant shortcomings:

- When a category overrides an inherited method, the method in the category can, as usual, invoke the inherited implementation via a message to `super`. However, if a category overrides a method that already existed in the category's class, there is no way to invoke the original implementation.
- A category cannot reliably override methods declared in another category of the same class.

This issue is of particular significance since many of the Cocoa classes are implemented using categories. A framework-defined method you try to override may itself have been implemented in a category, and so which implementation takes precedence is not defined.

- The very presence of some methods may cause behavior changes across all frameworks. For example, if you add an implementation of `windowWillClose:` to `NSObject`, this will cause all window delegates to respond to that method and may modify the behavior of all instances of `NSWindow` instances. This may cause mysterious changes in behavior and can lead to crashes.

Categories of the Root Class

A category can add methods to any class, including the root class. Methods added to `NSObject` become available to all classes that are linked to your code. While this can be useful at times, it can also be quite dangerous. Although it may seem that the modifications the category makes are well understood and of limited impact, inheritance gives them a wide scope. You may be making unintended changes to unseen classes; you may not know all the consequences of what you're doing. Moreover, others who are unaware of your changes won't understand what they're doing.

In addition, there are two other considerations to keep in mind when implementing methods for the root class:

- Messages to `super` are invalid (there is no superclass).
- Class objects can perform instance methods defined in the root class.

Normally, class objects can perform only class methods. But instance methods defined in the root class are a special case. They define an interface to the runtime system that all objects inherit. Class objects are full-fledged objects and need to share the same interface.

This feature means that you need to take into account the possibility that an instance method you define in a category of the `NSObject` class might be performed not only by instances but by class objects as well. For example, within the body of the method, `self` might mean a class object as well as an instance. See the `NSObject` class specification in the Foundation framework reference for more information on class access to root instance methods.

Extensions

Class extensions are like "anonymous" categories, except that the methods they declare must be implemented in the main `@implementation` block for the corresponding class.

It is common for a class to have a publicly declared API and to then have additional API declared privately for use solely by the class or the framework within which the class resides. You can declare such API in a category (or in more than one category) in a private header file or implementation file as described above. This works, but the compiler cannot verify that all declared methods are implemented.

For example, the compiler will compile without error the following declarations and implementation:

```
@interface MyObject : NSObject
{
```

```
    NSNumber *number;
}
- (NSNumber *)number;
@end

@interface MyObject (Setter)
- (void)setNumber:(NSNumber *)newNumber;
@end


@implementation MyObject

- (NSNumber *)number {
    return number;
}
@end
```

Note that there is no implementation of the `setNumber:` method. If it is invoked at runtime, this will generate an error.

Class extensions allow you to declare additional *required* API for a class in locations other than within the primary class `@interface` block, as illustrated in the following example:

```
@interface MyObject : NSObject
{
    NSNumber *number;
}
- (NSNumber *)number;
@end

@interface MyObject ()
- (void)setNumber:(NSNumber *)newNumber;
@end


@implementation MyObject

- (NSNumber *)number {
    return number;
}
- (void)setNumber:(NSNumber *)newNumber {
    number = newNumber;
}
@end
```

Notice that in this case:

- No name is given in the parentheses in the second `@interface` block;
- The implementation of the `setNumber:` method appears within the main `@implementation` block for the class.

The implementation of the `setNumber:` method *must* appear within the main `@implementation` block for the class (you cannot implement it in a category). If this is not the case, the compiler will emit a warning that it cannot find a method definition for `setNumber:`.

Associative References

You use associative references to simulate the addition of object instance variables to an existing class.

Associative references are only available in Mac OS X v10.6 and later.

Adding Storage Outside a Class Definition

Using associative references, you can add storage to an object without modifying the class declaration. This may be useful if you do not have access to the source code for the class, or if for binary-compatibility reasons you cannot alter the layout of the object.

Associations are based on a key, so for any object you can add as many associations as you want, each using a different key. An association can also ensure that the associated object remains valid for at least the lifetime of the source object (without the possibility of introducing uncollectible cycles in a garbage-collected environment).

Creating Associations

You use the Objective-C runtime function `objc_setAssociatedObject` to make an association between one object and another. The function takes four arguments: the source object, a key, the value, and an association policy constant. Of these, the key and the association policy merit further discussion.

- The key is a `void` pointer. The key for each association must be unique. A typical pattern is to use a `static` variable.
- The policy specifies whether the associated object is assigned, retained, or copied, and whether the association is be made atomically or non-atomically. This follows a similar pattern to the attributes of a declared property (see "Property Declaration Attributes" (page 68)). You specify the policy for the relationship using a constant (see `objc_AssociationPolicy`).

The following example shows how you can establish an association between an array and a string.

Listing 7-1 Establishing an association between an array and a string

```
static char overviewKey;

NSArray *array = [[NSArray alloc] initWithObjects:@"One", @"Two", @"Three",
nil];
// For the purposes of illustration, use initWithFormat: to ensure the string
can be deallocated
NSString *overview = [[NSString alloc] initWithFormat:@"%@", @"First three
numbers"];
```

```
objc_setAssociatedObject(array, &overviewKey, overview, OBJC_ASSOCIATION_RETAIN);

[overview release];
// (1) overview valid
[array release];
// (2) overview invalid
```

At point (1), the string `overview` is still valid because the `OBJC_ASSOCIATION_RETAIN` policy specifies that the array retains the associated object. When the array is deallocated, however (at point 2), `overview` is released and so in this case also deallocated. If you try to, for example, log the value of `overview`, you generate a runtime exception.

Retrieving Associated Objects

You retrieve an associated object using the Objective-C runtime function `objc_getAssociatedObject`. Continuing the example shown in Listing 7-1 (page 83), you could retrieve the overview from the array using the following line of code:

```
NSString *associatedObject = (NSString *)objc_getAssociatedObject(array,
&overviewKey);
```

Breaking Associations

To break an association, you typically use `objc_setAssociatedObject`, passing `nil` as the value.

Continuing the example shown in Listing 7-1 (page 83), you could break the association between the array and the string `overview` using the following line of code:

```
objc_setAssociatedObject(array, &overviewKey, nil, OBJC_ASSOCIATION_ASSIGN);
```

(Given that the associated object is being set to `nil`, the policy isn't actually important.)

To break *all* associations for an object, you can use `objc_removeAssociatedObjects`. In general, however, you are discouraged from using this since this breaks all associations for all clients. You only use this function if you need to restore an object to "pristine condition."

Complete Example

The following program combines the code samples from the preceding sections.

```
#import <Foundation/Foundation.h>
#import <objc/runtime.h>

int main (int argc, const char * argv[]) {
    NSAutoreleasePool * pool = [[NSAutoreleasePool alloc] init];

    static char overviewKey;
```

```
    NSArray *array = [[NSArray alloc] initWithObjects:@"One", @"Two", @"Three",
 nil];
    // For the purposes of illustration, use initWithFormat: to ensure we get
a
    // deallocatable string
    NSString *overview = [[NSString alloc] initWithFormat:@"%@", @"First three
 numbers"];

    objc_setAssociatedObject(array, &overviewKey, overview,
OBJC_ASSOCIATION_RETAIN);
    [overview release];

    NSString *associatedObject = (NSString *)objc_getAssociatedObject(array,
&overviewKey);
    NSLog(@"associatedObject: %@", associatedObject);

   objc_setAssociatedObject(array, &overviewKey, nil, OBJC_ASSOCIATION_ASSIGN);
    [array release];

    [pool drain];
    return 0;
}
```

Fast Enumeration

Fast enumeration is a language feature that allows you to efficiently and safely enumerate over the contents of a collection using a concise syntax.

The for…in Feature

Fast enumeration is a language feature that allows you to enumerate over the contents of a collection. The syntax is defined as follows:

```
for ( Type newVariable in expression ) { statements }
```

or

```
Type existingItem;
for ( existingItem in expression ) { statements }
```

In both cases, *expression* yields an object that conforms to the `NSFastEnumeration` protocol (see "Adopting Fast Enumeration" (page 87)). The iterating variable is set to each item in the returned object in turn, and the code defined by `statements` is executed. The iterating variable is set to `nil` when the loop ends by exhausting the source pool of objects. If the loop is terminated early, the iterating variable is left pointing to the last iteration item.

There are several advantages to using fast enumeration:

- The enumeration is considerably more efficient than, for example, using `NSEnumerator` directly.
- The syntax is concise.
- Enumeration is "safe"—the enumerator has a mutation guard so that if you attempt to modify the collection during enumeration, an exception is raised.

Since mutation of the object during iteration is forbidden, you can perform multiple enumerations concurrently.

Adopting Fast Enumeration

Any class whose instances provide access to a collection of other objects can adopt the `NSFastEnumeration` protocol. The Cocoa collection classes—`NSArray`, `NSDictionary`, and `NSSet`—adopt this protocol, as does `NSEnumerator`. It should be obvious that in the cases of `NSArray` and `NSSet` the enumeration is over their contents. For other classes, the corresponding documentation should make clear what property is iterated over—for example, `NSDictionary` and the Core Data class `NSManagedObjectModel` provide support for fast enumeration; `NSDictionary` enumerates its keys, and `NSManagedObjectModel` enumerates its entities.

Using Fast Enumeration

The following code example illustrates using fast enumeration with `NSArray` and `NSDictionary` objects.

```
NSArray *array = [NSArray arrayWithObjects:
        @"One", @"Two", @"Three", @"Four", nil];

for (NSString *element in array) {
    NSLog(@"element: %@", element);
}

NSDictionary *dictionary = [NSDictionary dictionaryWithObjectsAndKeys:
    @"quattuor", @"four", @"quinque", @"five", @"sex", @"six", nil];

NSString *key;
for (key in dictionary) {
    NSLog(@"English: %@, Latin: %@", key, [dictionary valueForKey:key]);
}
```

You can also use `NSEnumerator` objects with fast enumeration, as illustrated in the following example:

```
NSArray *array = [NSArray arrayWithObjects:
        @"One", @"Two", @"Three", @"Four", nil];

NSEnumerator *enumerator = [array reverseObjectEnumerator];
for (NSString *element in enumerator) {
    if ([element isEqualToString:@"Three"]) {
        break;
    }
}

NSString *next = [enumerator nextObject];
// next = "Two"
```

For collections or enumerators that have a well-defined order—such as `NSArray` or `NSEnumerator` instance derived from an array—the enumeration proceeds in that order, so simply counting iterations will give you the proper index into the collection if you need it.

```
NSArray *array = /* assume this exists */;
NSUInteger index = 0;

for (id element in array) {
    NSLog(@"Element at index %u is: %@", index, element);
    index++;
}
```

In other respects, the feature behaves like a standard `for` loop. You can use `break` to interrupt the iteration; and if you want to skip elements you can use a nested conditional statement as shown in the following example:

```
NSArray *array = /* assume this exists */;

for (id element in array) {
    if (/* some test for element */) {
        // statements that apply only to elements passing test
    }
```

```
}
```

If you want to skip the first element then process no more than five further elements, you could do so as shown in this example:

```
NSArray *array = /* assume this exists */;
NSUInteger index = 0;

for (id element in array) {
     if (index != 0) {
          NSLog(@"Element at index %u is: %@", index, element);
     }

     if (++index >= 6) {
          break;
     }
}
```

Enabling Static Behavior

This chapter explains how static typing works and discusses some other features of Objective-C, including ways to temporarily overcome its inherent dynamism.

Default Dynamic Behavior

By design, Objective-C objects are dynamic entities. As many decisions about them as possible are pushed from compile time to runtime:

- The memory for objects is **dynamically allocated** at runtime by class methods that create new instances.
- Objects are **dynamically typed**. In source code (at compile time), any object variable can be of type `id` no matter what the object's class is. The exact class of an `id` variable (and therefore its particular methods and data structure) isn't determined until the program runs.
- Messages and methods are **dynamically bound**, as described in "Dynamic Binding" (page 19). A runtime procedure matches the method selector in the message to a method implementation that "belongs to" the receiver.

These features give object-oriented programs a great deal of flexibility and power, but there's a price to pay. In particular, the compiler can't check the exact types (classes) of `id` variables. To permit better compile-time type checking, and to make code more self-documenting, Objective-C allows objects to be statically typed with a class name rather than generically typed as `id`. It also lets you turn some of its object-oriented features off in order to shift operations from runtime back to compile time.

Note: Messages are somewhat slower than function calls, typically incurring an insignificant amount of overhead compared to actual work performed. The exceptionally rare case where bypassing Objective-C's dynamism might be warranted can be proven by use of analysis tools like Shark or Instruments.

Static Typing

If a pointer to a class name is used in place of `id` in an object declaration,

```
Rectangle *thisObject;
```

the compiler restricts the value of the declared variable to be either an instance of the class named in the declaration or an instance of a class that inherits from the named class. In the example above, `thisObject` can only be a Rectangle of some kind.

Statically typed objects have the same internal data structures as objects declared to be `id`s. The type doesn't affect the object; it affects only the amount of information given to the compiler about the object and the amount of information available to those reading the source code.

Static typing also doesn't affect how the object is treated at runtime. Statically typed objects are dynamically allocated by the same class methods that create instances of type `id`. If Square is a subclass of Rectangle, the following code would still produce an object with all the instance variables of a Square, not just those of a Rectangle:

```
Rectangle *thisObject = [[Square alloc] init];
```

Messages sent to statically typed objects are dynamically bound, just as objects typed `id` are. The exact type of a statically typed receiver is still determined at runtime as part of the messaging process. A `display` message sent to `thisObject`

```
[thisObject display];
```

performs the version of the method defined in the Square class, not the one in its Rectangle superclass.

By giving the compiler more information about an object, static typing opens up possibilities that are absent for objects typed `id`:

- In certain situations, it allows for compile-time type checking.
- It can free objects from the restriction that identically named methods must have identical return and argument types.
- It permits you to use the structure pointer operator to directly access an object's instance variables.

The first two topics are discussed in the sections that follow. The third is covered in "Defining a Class" (page 35).

Type Checking

With the additional information provided by static typing, the compiler can deliver better type-checking services in two situations:

- When a message is sent to a statically typed receiver, the compiler can make sure the receiver can respond. A warning is issued if the receiver doesn't have access to the method named in the message.
- When a statically typed object is assigned to a statically typed variable, the compiler makes sure the types are compatible. A warning is issued if they're not.

An assignment can be made without warning, provided the class of the object being assigned is identical to, or inherits from, the class of the variable receiving the assignment. The following example illustrates this:

```
Shape     *aShape;
Rectangle *aRect;

aRect = [[Rectangle alloc] init];
aShape = aRect;
```

Here `aRect` can be assigned to `aShape` because a Rectangle is a kind of Shape—the Rectangle class inherits from Shape. However, if the roles of the two variables are reversed and `aShape` is assigned to `aRect`, the compiler generates a warning; not every Shape is a Rectangle. (For reference, see Figure 1-2 (page 25), which shows the class hierarchy including Shape and Rectangle.)

There's no check when the expression on either side of the assignment operator is an `id`. A statically typed object can be freely assigned to an `id`, or an `id` to a statically typed object. Because methods like `alloc` and `init` return `id`s, the compiler doesn't ensure that a compatible object is returned to a statically typed variable. The following code is error-prone, but is allowed nonetheless:

```
Rectangle *aRect;
aRect = [[Shape alloc] init];
```

Return and Argument Types

In general, methods in different classes that have the same selector (the same name) must also share the same return and argument types. This constraint is imposed by the compiler to allow dynamic binding. Because the class of a message receiver (and therefore class-specific details about the method it's asked to perform), can't be known at compile time, the compiler must treat all methods with the same name alike. When it prepares information on method return and argument types for the runtime system, it creates just one method description for each method selector.

However, when a message is sent to a statically typed object, the class of the receiver is known by the compiler. The compiler has access to class-specific information about the methods. Therefore, the message is freed from the restrictions on its return and argument types.

Static Typing to an Inherited Class

An instance can be statically typed to its own class or to any class that it inherits from. All instances, for example, can be statically typed as `NSObject`.

However, the compiler understands the class of a statically typed object only from the class name in the type designation, and it does its type checking accordingly. Typing an instance to an inherited class can therefore result in discrepancies between what the compiler thinks would happen at runtime and what actually happens.

For example, if you statically type a Rectangle instance as a Shape,

```
Shape *myRectangle = [[Rectangle alloc] init];
```

the compiler will treat it as a Shape. If you send the object a message to perform a Rectangle method,

```
BOOL solid = [myRectangle isFilled];
```

the compiler will complain. The `isFilled` method is defined in the Rectangle class, not in Shape.

However, if you send it a message to perform a method that the Shape class knows about,

```
[myRectangle display];
```

the compiler won't complain, even though Rectangle overrides the method. At runtime, Rectangle's version of the method is performed.

Similarly, suppose that the Upper class declares a `worry` method that returns a `double`,

```
- (double)worry;
```

and the Middle subclass of Upper overrides the method and declares a new return type:

```
- (int)worry;
```

If an instance is statically typed to the Upper class, the compiler will think that its `worry` method returns a `double`, and if an instance is typed to the Middle class, it will think that `worry` returns an `int`. Errors will obviously result if a Middle instance is typed to the Upper class. The compiler will inform the runtime system that a `worry` message sent to the object returns a `double`, but at runtime it actually returns an `int` and generates an error.

Static typing can free identically named methods from the restriction that they must have identical return and argument types, but it can do so reliably only if the methods are declared in different branches of the class hierarchy.

Selectors

In Objective-C, "selector" has two meanings. It can be used to refer simply to the name of a method when it's used in a source-code message to an object. It also, though, refers to the unique identifier that replaces the name when the source code is compiled. Compiled selectors are of type `SEL`. All methods with the same name have the same selector. You can use a selector to invoke a method on an object—this provides the basis for the implementation of the target-action design pattern in Cocoa.

Methods and Selectors

For efficiency, full ASCII names are not used as method selectors in compiled code. Instead, the compiler writes each method name into a table, then pairs the name with a unique identifier that represents the method at runtime. The runtime system makes sure each identifier is unique: No two selectors are the same, and all methods with the same name have the same selector.

SEL and @selector

Compiled selectors are assigned to a special type, `SEL`, to distinguish them from other data. Valid selectors are never `0`. You must let the system assign `SEL` identifiers to methods; it's futile to assign them arbitrarily.

The `@selector()` directive lets you refer to the compiled selector, rather than to the full method name. Here, the selector for `setWidth:height:` is assigned to the `setWidthHeight` variable:

```
SEL setWidthHeight;
setWidthHeight = @selector(setWidth:height:);
```

It's most efficient to assign values to `SEL` variables at compile time with the `@selector()` directive. However, in some cases, you may need to convert a character string to a selector at runtime. You can do this with the `NSSelectorFromString` function:

```
setWidthHeight = NSSelectorFromString(aBuffer);
```

Conversion in the opposite direction is also possible. The `NSStringFromSelector` function returns a method name for a selector:

```
NSString *method;
method = NSStringFromSelector(setWidthHeight);
```

Methods and Selectors

Compiled selectors identify method names, not method implementations. The `display` method for one class, for example, has the same selector as `display` methods defined in other classes. This is essential for polymorphism and dynamic binding; it lets you send the same message to receivers belonging to different classes. If there were one selector per method implementation, a message would be no different than a function call.

A class method and an instance method with the same name are assigned the same selector. However, because of their separate domains, there's no confusion between the two. A class could define a `display` class method in addition to a `display` instance method.

Method Return and Argument Types

The messaging routine has access to method implementations only through selectors, so it treats all methods with the same selector alike. It discovers the return type of a method, and the data types of its arguments, from the selector. Therefore, except for messages sent to statically typed receivers, dynamic binding requires all implementations of identically named methods to have the same return type and the same argument types. (Statically typed receivers are an exception to this rule, since the compiler can learn about the method implementation from the class type.)

Although identically named class methods and instance methods are represented by the same selector, they can have different argument and return types.

Varying the Message at Runtime

The `performSelector:`, `performSelector:withObject:`, and `performSelector:withObject:withObject:` methods, defined in the `NSObject` protocol, take `SEL` identifiers as their initial arguments. All three methods map directly into the messaging function. For example,

```
[friend performSelector:@selector(gossipAbout:)
    withObject:aNeighbor];
```

is equivalent to:

```
[friend gossipAbout:aNeighbor];
```

These methods make it possible to vary a message at runtime, just as it's possible to vary the object that receives the message. Variable names can be used in both halves of a message expression:

```
id   helper = getTheReceiver();
SEL  request = getTheSelector();
[helper performSelector:request];
```

In this example, the receiver (`helper`) is chosen at runtime (by the fictitious `getTheReceiver` function), and the method the receiver is asked to perform (`request`) is also determined at runtime (by the equally fictitious `getTheSelector` function).

Note: `performSelector:` and its companion methods return an `id`. If the method that's performed returns a different type, it should be cast to the proper type. (However, casting doesn't work for all types; the method should return a pointer or a type compatible with a pointer.)

The Target-Action Design Pattern

In its treatment of user-interface controls, the Application Kit makes good use of the ability to vary both the receiver and the message.

`NSControl` objects are graphical devices that can be used to give instructions to an application. Most resemble real-world control devices such as buttons, switches, knobs, text fields, dials, menu items, and the like. In software, these devices stand between the application and the user. They interpret events coming from hardware devices like the keyboard and mouse and translate them into application-specific instructions. For example, a button labeled "Find" would translate a mouse click into an instruction for the application to start searching for something.

The Application Kit defines a template for creating control devices and defines a few "off-the-shelf" devices of its own. For example, the `NSButtonCell` class defines an object that you can assign to an `NSMatrix` instance and initialize with a size, a label, a picture, a font, and a keyboard alternative. When the user clicks the button (or uses the keyboard alternative), the `NSButtonCell` object sends a message instructing the application to do something. To do this, an `NSButtonCell` object must be initialized not just with an image, a size, and a label, but with directions on what message to send and who to send it to. Accordingly, an `NSButtonCell` instance can be initialized for an action message, the method selector it should use in the message it sends, and a target, the object that should receive the message.

```
[myButtonCell setAction:@selector(reapTheWind:)];
[myButtonCell setTarget:anObject];
```

The button cell sends the message using `NSObject`'s `performSelector:withObject:` method. All action messages take a single argument, the `id` of the control device sending the message.

If Objective-C didn't allow the message to be varied, all `NSButtonCell` objects would have to send the same message; the name of the method would be frozen in the `NSButtonCell` source code. Instead of simply implementing a mechanism for translating user actions into action messages, button cells and other controls would have to constrain the content of the message. This would make it difficult for any object to respond to more than one button cell. There would either have to be one target for each button, or the target object would have to discover which button the message came from and act accordingly. Each time you rearranged the user interface, you would also have to re-implement the method that responds to the action message. This would be an unnecessary complication that Objective-C happily avoids.

Avoiding Messaging Errors

If an object receives a message to perform a method that isn't in its repertoire, an error results. It's the same sort of error as calling a nonexistent function. But because messaging occurs at runtime, the error often isn't evident until the program executes.

It's relatively easy to avoid this error when the message selector is constant and the class of the receiving object is known. As you write your programs, you can make sure that the receiver is able to respond. If the receiver is statically typed, the compiler performs this test for you.

However, if the message selector or the class of the receiver varies, it may be necessary to postpone this test until runtime. The `respondsToSelector:` method, defined in the `NSObject` class, determines whether a receiver can respond to a message. It takes the method selector as an argument and returns whether the receiver has access to a method matching the selector:

```
if ( [anObject respondsToSelector:@selector(setOrigin::)] )
    [anObject setOrigin:0.0 :0.0];
else
    fprintf(stderr, "%s can't be placed\n",
        [NSStringFromClass([anObject class]) UTF8String]);
```

The `respondsToSelector:` test is especially important when sending messages to objects that you don't have control over at compile time. For example, if you write code that sends a message to an object represented by a variable that others can set, you should make sure the receiver implements a method that can respond to the message.

Note: An object can also arrange to have the messages it receives forwarded to other objects if it can't respond to them directly itself. In that case, it appears that the object can handle the message, even though the object responds to the message indirectly by assigning it to another object. See Message Forwarding in *Objective-C Runtime Programming Guide* for more information.

Exception Handling

The Objective-C language has an exception-handling syntax similar to that of Java and C++. Coupled with the use of the `NSException`, `NSError`, or custom classes, you can add robust error-handling to your programs. This article provides a summary of exception syntax and handling; for more details, see *Exception Programming Topics*.

Enabling Exception-Handling

Using GNU Compiler Collection (GCC) version 3.3 and later, Objective-C provides language-level support for exception handling. To turn on support for these features, use the `-fobjc-exceptions` switch of the GNU Compiler Collection (GCC) version 3.3 and later. (Note that this renders the application runnable only in Mac OS X v10.3 and later because runtime support for exception handling and synchronization is not present in earlier versions of the software.)

Exception Handling

An exception is a special condition that interrupts the normal flow of program execution. There are a variety of reasons why an exception may be generated (exceptions are typically said to be **raised** or **thrown**), by hardware as well as software. Examples include arithmetical errors such as division by zero, underflow or overflow, calling undefined instructions (such as attempting to invoke an unimplemented method), and attempting to access a collection element out of bounds.

Objective-C's exception support revolves around four compiler directives: `@try`, `@catch`, `@throw`, and `@finally`:

- Code that can potentially throw an exception is enclosed in a `@try` block.
- A `@catch()` block contains exception-handling logic for exceptions thrown in a `@try` block. You can have multiple `@catch()` blocks to catch different types of exception. (This is illustrated in "Catching Different Types of Exception" (page 100).)
- A `@finally` block contains code that must be executed whether an exception is thrown or not.
- You use the `@throw` directive to throw an exception, which is essentially an Objective-C object. You typically use an `NSException` object, but are not required to.

The example below depicts a simple exception-handling algorithm:

```
Cup *cup = [[Cup alloc] init];

@try {
    [cup fill];
}
```

```
@catch (NSException *exception) {
    NSLog(@"main: Caught %@: %@", [exception name], [exception reason]);
}
@finally {
    [cup release];
}
```

Catching Different Types of Exception

To catch an exception thrown in a `@try` block, use one or more `@catch()`blocks following the `@try` block. The `@catch()` blocks should be ordered from most-specific to the least-specific. That way you can tailor the processing of exceptions as groups, as shown in Listing 11-1.

Listing 11-1 An exception handler

```
@try {
    ...
}
@catch (CustomException *ce) {  // 1
    ...
}
@catch (NSException *ne) {  // 2
    // Perform processing necessary at this level.
    ...

}
@catch (id ue) {
    ...
}
@finally {  // 3
    // Perform processing necessary whether an exception occurred  or not.
    ...
}
```

The following list describes the numbered code-lines:

1. Catches the most specific exception type.
2. Catches a more general exception type.
3. Performs any clean-up processing that must always be performed, whether exceptions were thrown or not.

Throwing Exceptions

To throw an exception you must instantiate an object with the appropriate information, such as the exception name and the reason it was thrown.

```
NSException *exception = [NSException exceptionWithName:@"HotTeaException"
                            reason:@"The tea is too hot"  userInfo:nil];
```

```
@throw exception;
```

Important: In many environments, use of exceptions is fairly commonplace. For example, you might throw an exception to signal that a routine could not execute normally—such as when a file is missing or data could not be parsed correctly. Exceptions are resource-intensive in Objective-C. You should not use exceptions for general flow-control, or simply to signify errors. Instead you should use the return value of a method or function to indicate that an error has occurred, and provide information about the problem in an error object. For more information, see *Error Handling Programming Guide*.

Inside a `@catch()` block, you can re-throw the caught exception using the `@throw` directive without an argument. This can help make your code more readable.

You are not limited to throwing `NSException` objects. You can throw any Objective-C object as an exception object. The `NSException` class provides methods that help in exception processing, but you can implement your own if you so desire. You can also subclass `NSException` to implement specialized types of exceptions, such as file-system exceptions or communications exceptions.

Threading

Objective-C provides support for thread synchronization and exception handling, which are explained in this article and "Exception Handling" (page 99). To turn on support for these features, use the `-fobjc-exceptions` switch of the GNU Compiler Collection (GCC) version 3.3 and later.

Note: Using either of these features in a program, renders the application runnable only in Mac OS X v10.3 and later because runtime support for exception handling and synchronization is not present in earlier versions of the software.

Synchronizing Thread Execution

Objective-C supports multithreading in applications. This means that two threads can try to modify the same object at the same time, a situation that can cause serious problems in a program. To protect sections of code from being executed by more than one thread at a time, Objective-C provides the `@synchronized()` directive.

The `@synchronized()`directive locks a section of code for use by a single thread. Other threads are blocked until the thread exits the protected code; that is, when execution continues past the last statement in the `@synchronized()` block.

The `@synchronized()` directive takes as its only argument any Objective-C object, including `self`. This object is known as a *mutual exclusion* semaphore or *mutex*. It allows a thread to lock a section of code to prevent its use by other threads. You should use separate semaphores to protect different critical sections of a program. It's safest to create all the mutual exclusion objects before the application becomes multithreaded to avoid race conditions.

Listing 12-1 shows an example of code that uses `self` as the mutex to synchronize access to the instance methods of the current object. You can take a similar approach to synchronize the class methods of the associated class, using the Class object instead of `self`. In the latter case, of course, only one thread at a time is allowed to execute a class method because there is only one class object that is shared by all callers.

Listing 12-1 Locking a method using self

```
- (void)criticalMethod
{
    @synchronized(self) {
        // Critical code.
        ...
    }
}
```

Listing 12-2 shows a general approach. Before executing a critical process, the code obtains a semaphore from the Account class and uses it to lock the critical section. The Account class could create the semaphore in its `initialize` method.

Listing 12-2 Locking a method using a custom semaphore

```
Account *account = [Account accountFromString:[accountField stringValue]];

// Get the semaphore.
id accountSemaphore = [Account semaphore];

@synchronized(accountSemaphore) {
    // Critical code.
    ...
}
```

The Objective-C synchronization feature supports recursive and reentrant code. A thread can use a single semaphore several times in a recursive manner; other threads are blocked from using it until the thread releases all the locks obtained with it; that is, every `@synchronized()` block is exited normally or through an exception.

When code in an `@synchronized()` block throws an exception, the Objective-C runtime catches the exception, releases the semaphore (so that the protected code can be executed by other threads), and re-throws the exception to the next exception handler.

Remote Messaging

Like most other programming languages, Objective-C was initially designed for programs that are executed as a single process in a single address space.

Nevertheless, the object-oriented model, where communication takes place between relatively self-contained units through messages that are resolved at runtime, would seem well suited for interprocess communication as well. It's not hard to imagine Objective-C messages between objects that reside in different address spaces (that is, in different tasks) or in different threads of execution of the same task.

For example, in a typical server-client interaction, the client task might send its requests to a designated object in the server, and the server might target specific client objects for the notifications and other information it sends. Or imagine an interactive application that needs to do a good deal of computation to carry out a user command. It could simply display a dialog telling the user to wait while it was busy, or it could isolate the processing work in a subordinate task, leaving the main part of the application free to accept user input. Objects in the two tasks would communicate through Objective-C messages.

Distributed Objects

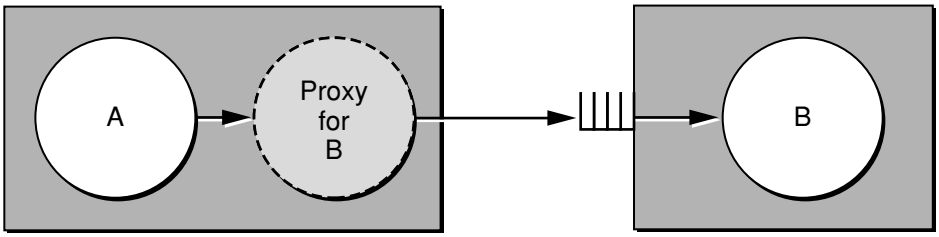
Remote messaging in Objective-C requires a runtime system that can establish connections between objects in different address spaces, recognize when a message is intended for an object in a remote address space, and transfer data from one address space to another. It must also mediate between the separate schedules of the two tasks; it has to hold messages until their remote receivers are free to respond to them.

Cocoa includes a **distributed objects** architecture that is essentially this kind of extension to the runtime system. Using distributed objects, you can send Objective-C messages to objects in other tasks or have messages executed in other threads of the same task. (When remote messages are sent between two threads of the same task, the threads are treated exactly like threads in different tasks.) Note that Cocoa's distributed objects system is built on top of the runtime system; it doesn't alter the fundamental behavior of your Cocoa objects.

To send a remote message, an application must first establish a connection with the remote receiver. Establishing the connection gives the application a proxy for the remote object in its own address space. It then communicates with the remote object through the proxy. The proxy assumes the identity of the remote object; it has no identity of its own. The application is able to regard the proxy as if it were the remote object; for most purposes, it is the remote object.

Remote messaging is illustrated in Figure 13-1 (page 106), where object A communicates with object B through a proxy, and messages for B wait in a queue until B is ready to respond to them:

Figure 13-1 Remote Messages



The sender and receiver are in different tasks and are scheduled independently of each other. So there's no guarantee that the receiver is free to accept a message when the sender is ready to send it. Therefore, arriving messages are placed in a queue and retrieved at the convenience of the receiving application.

A proxy doesn't act on behalf of the remote object or need access to its class. It isn't a copy of the object, but a lightweight substitute for it. In a sense, it's transparent; it simply passes the messages it receives on to the remote receiver and manages the interprocess communication. Its main function is to provide a local address for an object that wouldn't otherwise have one. A proxy isn't fully transparent, however. For instance, a proxy doesn't allow you to directly set and get an object's instance variables.

A remote receiver is typically anonymous. Its class is hidden inside the remote application. The sending application doesn't need to know how that application is designed or what classes it uses. It doesn't need to use the same classes itself. All it needs to know is what messages the remote object responds to.

Because of this, an object that's designated to receive remote messages advertises its interface in a formal protocol. Both the sending and the receiving application declare the protocol—they both import the same protocol declaration. The receiving application declares it because the remote object must conform to the protocol. The sending application declares it to inform the compiler about the messages it sends and because it may use the `conformsToProtocol:` method and the `@protocol()` directive to test the remote receiver. The sending application doesn't have to implement any of the methods in the protocol; it declares the protocol only because it initiates messages to the remote receiver.

The distributed objects architecture, including the `NSProxy` and `NSConnection` classes, is documented in the Foundation framework reference and *Distributed Objects Programming Topics*.

Language Support

Remote messaging raises not only a number of intriguing possibilities for program design, it also raises some interesting issues for the Objective-C language. Most of the issues are related to the efficiency of remote messaging and the degree of separation that the two tasks should maintain while they're communicating with each other.

So that programmers can give explicit instructions about the intent of a remote message, Objective-C defines six type qualifiers that can be used when declaring methods inside a formal protocol:

```
oneway
in
out
inout
bycopy
```

`byref`

These modifiers are restricted to formal protocols; they can't be used inside class and category declarations. However, if a class or category adopts a protocol, its implementation of the protocol methods can use the same modifiers that are used to declare the methods.

The following sections explain how these modifiers are used.

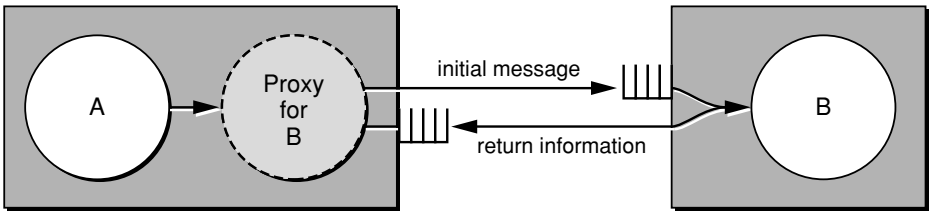
Synchronous and Asynchronous Messages

Consider first a method with just a simple return value:

```
- (BOOL)canDance;
```

When a `canDance` message is sent to a receiver in the same application, the method is invoked and the return value provided directly to the sender. But when the receiver is in a remote application, two underlying messages are required—one message to get the remote object to invoke the method, and the other message to send back the result of the remote calculation. This is illustrated in the figure below:

Figure 13-2 Round-Trip Message



Most remote messages are, at bottom, two-way (or "round trip") remote procedure calls (RPCs) like this one. The sending application waits for the receiving application to invoke the method, complete its processing, and send back an indication that it has finished, along with any return information requested. Waiting for the receiver to finish, even if no information is returned, has the advantage of coordinating the two communicating applications, of keeping them both "in sync." For this reason, round-trip messages are often called **synchronous**. Synchronous messages are the default.

However, it's not always necessary or a good idea to wait for a reply. Sometimes it's sufficient simply to dispatch the remote message and return, allowing the receiver to get to the task when it can. In the meantime, the sender can go on to other things. Objective-C provides a return type modifier, `oneway`, to indicate that a method is used only for **asynchronous** messages:

```
- (oneway void)waltzAtWill;
```

Although `oneway` is a type qualifier (like `const`) and can be used in combination with a specific type name, such as `oneway float` or `oneway id`, the only such combination that makes any sense is `oneway void`. An asynchronous message can't have a valid return value.

Pointer Arguments

Next, consider methods that take pointer arguments. A pointer can be used to pass information to the receiver by reference. When invoked, the method looks at what's stored in the address it's passed.

```
- setTune:(struct tune *)aSong
{
    tune = *aSong;
    ...
}
```

The same sort of argument can also be used to return information by reference. The method uses the pointer to find where it should place information requested in the message.

```
- getTune:(struct tune *)theSong
{
    ...
    *theSong = tune;
}
```

The way the pointer is used makes a difference in how the remote message is carried out. In neither case can the pointer simply be passed to the remote object unchanged; it points to a memory location in the sender's address space and would not be meaningful in the address space of the remote receiver. The runtime system for remote messaging must make some adjustments behind the scenes.

If the argument is used to pass information by reference, the runtime system must dereference the pointer, ship the value it points to over to the remote application, store the value in an address local to that application, and pass that address to the remote receiver.

If, on the other hand, the pointer is used to return information by reference, the value it points to doesn't have to be sent to the other application. Instead, a value from the other application must be sent back and written into the location indicated by the pointer.

In the first case, information is passed on the first leg of the round trip. In the second case, information is returned on the second leg of the round trip. Because these cases result in very different actions on the part of the runtime system for remote messaging, Objective-C provides type modifiers that can clarify the programmer's intention:

- The type modifier `in` indicates that information is being passed in a message:

  ```
  - setTune:(in struct tune *)aSong;
  ```

- The modifier `out` indicates that an argument is being used to return information by reference:

  ```
  - getTune:(out struct tune *)theSong;
  ```

- A third modifier, `inout`, indicates that an argument is used both to provide information and to get information back:

  ```
  - adjustTune:(inout struct tune *)aSong;
  ```

The Cocoa distributed objects system takes `inout` to be the default modifier for all pointer arguments except those declared `const`, for which `in` is the default. `inout` is the safest assumption but also the most time-consuming since it requires passing information in both directions. The only modifier that makes sense for arguments passed by value (non-pointers) is `in`. While `in` can be used with any kind of argument, `out` and `inout` make sense only for pointers.

In C, pointers are sometimes used to represent composite values. For example, a string is represented as a character pointer (`char *`). Although in notation and implementation there's a level of indirection here, in concept there's not. Conceptually, a string is an entity in and of itself, not a pointer to something else.

In cases like this, the distributed objects system automatically dereferences the pointer and passes whatever it points to as if by value. Therefore, the `out` and `inout` modifiers make no sense with simple character pointers. It takes an additional level of indirection in a remote message to pass or return a string by reference:

```
- getTuneTitle:(out char **)theTitle;
```

The same is true of objects:

```
- adjustRectangle:(inout Rectangle **)theRect;
```

These conventions are enforced at runtime, not by the compiler.

Proxies and Copies

Finally, consider a method that takes an object as an argument:

```
- danceWith:(id)aPartner;
```

A `danceWith:` message passes an object `id` to the receiver. If the sender and the receiver are in the same application, they would both be able to refer to the same *`aPartner`* object.

This is true even if the receiver is in a remote application, except that the receiver needs to refer to the object through a proxy (since the object isn't in its address space). The pointer that `danceWith:` delivers to a remote receiver is actually a pointer to the proxy. Messages sent to the proxy would be passed across the connection to the real object and any return information would be passed back to the remote application.

There are times when proxies may be unnecessarily inefficient, when it's better to send a copy of the object to the remote process so that it can interact with it directly in its own address space. To give programmers a way to indicate that this is intended, Objective-C provides a `bycopy` type modifier:

```
- danceWith:(bycopy id)aClone;
```

`bycopy` can also be used for return values:

```
- (bycopy)dancer;
```

It can similarly be used with `out` to indicate that an object returned by reference should be copied rather than delivered in the form of a proxy:

```
- getDancer:(bycopy out id *)theDancer;
```

Note: When a copy of an object is passed to another application, it cannot be anonymous. The application that receives the object must have the class of the object loaded in its address space.

`bycopy` makes so much sense for certain classes—classes that are intended to contain a collection of other objects, for instance—that often these classes are written so that a copy is sent to a remote receiver, instead of the usual reference. You can override this behavior with `byref`, however, thereby specifying that objects passed to a method or objects returned from a method should be passed or returned by reference. Since passing by reference is the default behavior for the vast majority of Objective-C objects, you will rarely, if ever, make use of the `byref` keyword.

The only type that it makes sense for `bycopy` or `byref` to modify is an object, whether dynamically typed `id` or statically typed by a class name.

Although `bycopy` and `byref` can't be used inside class and category declarations, they can be used within formal protocols. For instance, you could write a formal protocol `foo` as follows:

```
@Protocol foo
- (bycopy)array;
@end
```

A class or category can then adopt your protocol `foo`. This allows you to construct protocols so that they provide "hints" as to how objects should be passed and returned by the methods described by the protocol.

Using C++ With Objective-C

Apple's Objective-C compiler allows you to freely mix C++ and Objective-C code in the same source file. This Objective-C/C++ language hybrid is called Objective-C++. With it you can make use of existing C++ libraries from your Objective-C applications.

Mixing Objective-C and C++ Language Features

In Objective-C++, you can call methods from either language in C++ code and in Objective-C methods. Pointers to objects in either language are just pointers, and as such can be used anywhere. For example, you can include pointers to Objective-C objects as data members of C++ classes, and you can include pointers to C++ objects as instance variables of Objective-C classes. Listing 14-1 illustrates this.

Note: Xcode requires that file names have a ".mm" extension for the Objective-C++ extensions to be enabled by the compiler.

Listing 14-1 Using C++ and Objective-C instances as instance variables

```
/* Hello.mm
 * Compile with: g++ -x objective-c++ -framework Foundation Hello.mm  -o hello
 */

#import <Foundation/Foundation.h>
class Hello {
    private:
        id greeting_text;  // holds an NSString
    public:
        Hello() {
            greeting_text = @"Hello, world!";
        }
        Hello(const char* initial_greeting_text) {
            greeting_text = [[NSString alloc]
initWithUTF8String:initial_greeting_text];
        }
        void say_hello() {
            printf("%s\n", [greeting_text UTF8String]);
        }
};

@interface Greeting : NSObject {
    @private
        Hello *hello;
}
- (id)init;
- (void)dealloc;
- (void)sayGreeting;
- (void)sayGreeting:(Hello*)greeting;
```

```
@end

@implementation Greeting
- (id)init {
    if (self = [super init]) {
        hello = new Hello();
    }
    return self;
}
- (void)dealloc {
    delete hello;
    [super dealloc];
}
- (void)sayGreeting {
    hello->say_hello();
}
- (void)sayGreeting:(Hello*)greeting {
    greeting->say_hello();
}
@end

int main() {
    NSAutoreleasePool *pool = [[NSAutoreleasePool alloc] init];

    Greeting *greeting = [[Greeting alloc] init];
    [greeting sayGreeting];                         // > Hello,  world!

    Hello *hello = new Hello("Bonjour, monde!");
    [greeting sayGreeting:hello];                   // > Bonjour,  monde!

    delete hello;
    [greeting release];
    [pool release];
    return 0;
}
```

As you can declare C structs in Objective-C interfaces, you can also declare C++ classes in Objective-C interfaces. As with C structs, C++ classes defined within an Objective-C interface are globally-scoped, not nested within the Objective-C class. (This is consistent with the way in which standard C—though not C++—promotes nested struct definitions to file scope.)

To allow you to conditionalize your code based on the language variant, the Objective-C++ compiler defines both the `__cplusplus` and the `__OBJC__` preprocessor constants, as specified by (respectively) the C++ and Objective-C language standards.

As previously noted, Objective-C++ does not allow you to inherit C++ classes from Objective-C objects, nor does it allow you to inherit Objective-C classes from C++ objects.

```
class Base { /* ... */ };
@interface ObjCClass: Base ... @end // ERROR!
class Derived: public ObjCClass ... // ERROR!
```

Unlike Objective-C, objects in C++ are statically typed, with runtime polymorphism available as an exceptional case. The object models of the two languages are thus not directly compatible. More fundamentally, the layout of Objective-C and C++ objects in memory is mutually incompatible, meaning that it is generally impossible to create an object instance that would be valid from the perspective of both languages. Hence, the two type hierarchies cannot be intermixed.

You can declare a C++ class within an Objective-C class declaration. The compiler treats such classes as having been declared in the global namespace, as follows:

```
@interface Foo {
 class Bar { ... } // OK
}
@end

Bar *barPtr; // OK
```

Objective-C allows C structures (whether declared inside of an Objective-C declaration or not) to be used as instance variables.

```
@interface Foo {
   struct CStruct { ... };
   struct CStruct bigIvar; // OK
} ... @end
```

On Mac OS X 10.4 and later, if you set the `fobjc-call-cxx-cdtors` compiler flag, you can use instances of C++ classes containing virtual functions and nontrivial user-defined zero-argument constructors and destructors as instance variables. (The `fobjc-call-cxx-cdtors` compiler flag is set by default in gcc-4.2.) Constructors are invoked in the `alloc` method (specifically, inside `class_createInstance`), in declaration order immediately after the Objective-C object of which they are a member is allocated. The constructor used is the "public no-argument in-place constructor." Destructors are invoked in the `dealloc` method (specifically, inside `object_dispose`), in reverse declaration order immediately before the Objective-C object of which they are a member is deallocated.

Mac OS X v10.3 and earlier: The following cautions apply only to Mac OS X v10.3 and earlier.

Objective-C++ similarly strives to allow C++ class instances to serve as instance variables. This is possible as long as the C++ class in question (along with all of its superclasses) does not have any virtual member functions defined. If any virtual member functions are present, the C++ class may not serve as an Objective-C instance variable.

```
#import <Cocoa/Cocoa.h>

struct Class0 { void foo(); };
struct Class1 { virtual void foo(); };
struct Class2 { Class2(int i, int j); };

@interface Foo : NSObject {
    Class0 class0;      // OK
    Class1 class1;      // ERROR!
    Class1 *ptr;        // OK—call 'ptr = new Class1()' from Foo's init,
                        // 'delete ptr' from Foo's dealloc
    Class2 class2;      // WARNING - constructor not called!
...
@end
```

C++ requires each instance of a class containing virtual functions to contain a suitable virtual function table pointer. However, the Objective-C runtime cannot initialize the virtual function table pointer, because it is not familiar with the C++ object model. Similarly, the Objective-C runtime cannot dispatch calls to C++ constructors or destructors for those objects. If a C++ class has any user-defined constructors or destructors, they are not called. The compiler emits a warning in such cases.

Objective-C does not have a notion of nested namespaces. You cannot declare Objective-C classes within C++ namespaces, nor can you declare namespaces within Objective-C classes.

Objective-C classes, protocols, and categories cannot be declared inside a C++ template, nor can a C++ template be declared inside the scope of an Objective-C interface, protocol, or category.

However, Objective-C classes may serve as C++ template parameters. C++ template parameters can also be used as receivers or parameters (though not as selectors) in Objective-C message expressions.

C++ Lexical Ambiguities and Conflicts

There are a few identifiers that are defined in the Objective-C header files that every Objective-C program must include. These identifiers are `id`, `Class`, `SEL`, `IMP`, and `BOOL`.

Inside an Objective-C method, the compiler pre-declares the identifiers `self` and `super`, similarly to the keyword `this` in C++. However, unlike the C++ `this` keyword, `self` and `super` are context-sensitive; they may be used as ordinary identifiers outside of Objective-C methods.

In the parameter list of methods within a protocol, there are five more context-sensitive keywords (`oneway`, `in`, `out`, `inout`, and `bycopy`). These are not keywords in any other contexts.

From an Objective-C programmer's point of view, C++ adds quite a few new keywords. You can still use C++ keywords as a part of an Objective-C selector, so the impact isn't too severe, but you cannot use them for naming Objective-C classes or instance variables. For example, even though `class` is a C++ keyword, you can still use the `NSObject` method `class`:

```
[foo class]; // OK
```

However, because it is a keyword, you cannot use `class` as the name of a variable:

```
NSObject *class; // Error
```

In Objective-C, the names for classes and categories live in separate namespaces. That is, both `@interface foo` and `@interface(foo)` can exist in the same source code. In Objective-C++, you can also have a category whose name matches that of a C++ class or structure.

Protocol and template specifiers use the same syntax for different purposes:

```
id<someProtocolName> foo;
TemplateType<SomeTypeName> bar;
```

To avoid this ambiguity, the compiler doesn't permit `id` to be used as a template name.

Finally, there is a lexical ambiguity in C++ when a label is followed by an expression that mentions a global name, as in:

```
label: ::global_name = 3;
```

The space after the first colon is required. Objective-C++ adds a similar case, which also requires a space:

```
receiver selector: ::global_c++_name;
```

Limitations

Objective-C++ does not add C++ features to Objective-C classes, nor does it add Objective-C features to C++ classes. For example, you cannot use Objective-C syntax to call a C++ object, you cannot add constructors or destructors to an Objective-C object, and you cannot use the keywords `this` and `self` interchangeably. The class hierarchies are separate; a C++ class cannot inherit from an Objective-C class, and an Objective-C class cannot inherit from a C++ class. In addition, multi-language exception handling is not supported. That is, an exception thrown in Objective-C code cannot be caught in C++ code and, conversely, an exception thrown in C++ code cannot be caught in Objective-C code. For more information on exceptions in Objective-C, see "Exception Handling" (page 99).

Language Summary

Objective-C adds a small number of constructs to the C language and defines a handful of conventions for effectively interacting with the runtime system. This appendix lists all the additions to the language but doesn't go into great detail. For more information, see the other chapters in this document. .

Messages

Message expressions are enclosed in square brackets:

```
[receiver message]
```

The receiver can be:

- A variable or expression that evaluates to an object (including the variable `self`)
- A class name (indicating the class object)
- `super` (indicating an alternative search for the method implementation)

The **message** is the name of a method plus any arguments passed to it.

Defined Types

The principal types used in Objective-C are defined in `objc/objc.h`. They are:

Type	Definition
`id`	An object (a pointer to its data structure).
`Class`	A class object (a pointer to the class data structure).
`SEL`	A selector, a compiler-assigned code that identifies a method name.
`IMP`	A pointer to a method implementation that returns an `id`.
`BOOL`	A Boolean value, either `YES` or `NO`. Note that the type of `BOOL` is `char`.

`id` can be used to type any kind of object, class, or instance. In addition, class names can be used as type names to statically type instances of a class. A statically typed instance is declared to be a pointer to its class or to any class it inherits from.

The `objc.h` header file also defines these useful terms:

Type	Definition
`nil`	A null object pointer, `(id)0`.
`Nil`	A null class pointer, `(Class)0`.
`NO`	A boolean false value, `(BOOL)0`.
`YES`	A boolean true value, `(BOOL)1`.

Preprocessor Directives

The preprocessor understands these special notations:

Notation	Definition
`#import`	Imports a header file. This directive is identical to `#include`, except that it doesn't include the same file more than once.
`//`	Begins a comment that continues to the end of the line.

Compiler Directives

Directives to the compiler begin with "@". The following directives are used to declare and define classes, categories, and protocols:

Directive	Definition
`@interface`	Begins the declaration of a class or category interface.
`@implementation`	Begins the definition of a class or category.
`@protocol`	Begins the declaration of a formal protocol.
`@end`	Ends the declaration/definition of a class, category, or protocol.

The following mutually exclusive directives specify the visibility of instance variables:

Directive	Definition
`@private`	Limits the scope of an instance variable to the class that declares it.
`@protected`	Limits instance variable scope to declaring and inheriting classes.
`@public`	Removes restrictions on the scope of instance variables.

The default is `@protected`.

These directives support exception handling:

Directive	Definition
`@try`	Defines a block within which exceptions can be thrown.
`@throw`	Throws an exception object.
`@catch()`	Catches an exception thrown within the preceding `@try` block.
`@finally`	Defines a block of code that is executed whether exceptions were thrown or not in a preceding `@try` block.

The following directives support the declared properties feature (see "Declared Properties" (page 67)):

Directive	Definition
`@property`	Begins the declaration of a declared property.
`@synthesize`	Requests that, for the properties whose names follow, the compiler generate accessor methods for which there are no custom implementations.
`@dynamic`	Instructs the compiler not to generate a warning if it cannot find implementations of accessor methods associated with the properties whose names follow.

In addition, there are directives for these particular purposes:

Directive	Definition
`@class`	Declares the names of classes defined elsewhere.
`@selector(method_name)`	Returns the compiled selector that identifies *method_name*.
`@protocol(protocol_name)`	Returns the *protocol_name* protocol (an instance of the Protocol class). (`@protocol` is also valid without (*protocol_name*) for forward declarations.)
`@encode(type_spec)`	Yields a character string that encodes the type structure of *type_spec*.
`@"string"`	Defines a constant `NSString` object in the current module and initializes the object with the specified string. On Mac OS X v10.4 and earlier, the string must be 7-bit ASCII-encoded. On Mac OS X v10.5 and later (with Xcode 3.0 and later), you can also use UTF-16 encoded strings. (The runtime from Mac OS X v10.2 and later supports UTF-16 encoded strings, so if you use Mac OS X v10.5 to compile an application for Mac OS X v10.2 and later, you can use UTF-16 encoded strings.)

Directive	Definition
`@"string1" @"string2" ... @"stringN"`	Defines a constant `NSString` object in the current module. The string created is the result of concatenating the strings specified in the two directives.
`@synchronized()`	Defines a block of code that must be executed only by one thread at a time.

Classes

A new class is declared with the `@interface` directive. The interface file for its superclass must be imported:

```
#import "ItsSuperclass.h"

@interface ClassName : ItsSuperclass < protocol_list >
{
    instance variable declarations
}
method declarations
@end
```

Everything but the compiler directives and class name is optional. If the colon and superclass name are omitted, the class is declared to be a new root class. If any protocols are listed, the header files where they're declared must also be imported.

A file containing a class definition imports its own interface:

```
#import "ClassName.h"

@implementation ClassName
method definitions
@end
```

Categories

A category is declared in much the same way as a class. The interface file that declares the class must be imported:

```
#import "ClassName.h"

@interface ClassName ( CategoryName ) < protocol list >
method declarations
@end
```

The protocol list and method declarations are optional. If any protocols are listed, the header files where they're declared must also be imported.

Like a class definition, a file containing a category definition imports its own interface:

```
#import "CategoryName.h"

@implementation ClassName ( CategoryName )
method definitions
@end
```

Formal Protocols

Formal protocols are declared using the `@protocol` directive:

```
@protocol ProtocolName < protocol list >
declarations of required methods
@optional
declarations of optional methods
@required
declarations of required methods
@end
```

The list of incorporated protocols and the method declarations are optional. The protocol must import the header files that declare any protocols it incorporates.

The `@optional` directive specifies that following methods are optional; the `@required` directive directive specifies that following methods must be implemented by a class that adopts the protocol. The default is `@required`.

You can create a forward reference to a protocol using the `@protocol` directive in the following manner:

```
@protocol ProtocolName;
```

Within source code, protocols are referred to using the similar `@protocol()` directive, where the parentheses enclose the protocol name.

Protocol names listed within angle brackets (<...>) are used to do three different things:

- In a protocol declaration, to incorporate other protocols (as shown earlier)
- In a class or category declaration, to adopt the protocol (as shown in "Classes" (page 120) and "Categories" (page 120))
- In a type specification, to limit the type to objects that conform to the protocol

Within protocol declarations, these type qualifiers support remote messaging:

Type Qualifier	Definition
`oneway`	The method is for asynchronous messages and has no valid return type.
`in`	The argument passes information to the remote receiver.
`out`	The argument gets information returned by reference.
`inout`	The argument both passes information and gets information.
`bycopy`	A copy of the object, not a proxy, should be passed or returned.

Type Qualifier	Definition
`byref`	A reference to the object, not a copy, should be passed or returned.

Method Declarations

The following conventions are used in method declarations:

- A "+" precedes declarations of class methods.
- A "-" precedes declarations of instance methods.
- Argument and return types are declared using the C syntax for type casting.
- Arguments are declared after colons (:), for example:

  ```
  - (void)setWidth:(int)newWidth height:(int)newHeight
  ```

  Typically, a label describing the argument precedes the colon—the following example is valid but is considered bad style:

  ```
  - (void)setWidthAndHeight:(int)newWidth :(int)newHeight
  ```

  Both labels and colons are considered part of the method name.
- The default return and argument type for methods is `id`, not `int` as it is for functions. (However, the modifier `unsigned` when used without a following type always means `unsigned int`.)

Method Implementations

Each method implementation is passed two hidden arguments:

- The receiving object (`self`).
- The selector for the method (`_cmd`).

Within the implementation, both `self` and `super` refer to the receiving object. `super` replaces `self` as the receiver of a message to indicate that only methods inherited by the implementation should be performed in response to the message.

Methods with no other valid return typically return `void`.

Deprecation Syntax

Syntax is provided to mark methods as deprecated:

```
@interface SomeClass
-method __attribute__((deprecated));
```

```
@end
```

or:

```
#include <AvailabilityMacros.h>
@interface SomeClass
-method DEPRECATED_ATTRIBUTE;  // or some other deployment-target-specific macro
@end
```

This syntax is available only in Objective-C 2.0 and later.

Naming Conventions

The names of files that contain Objective-C source code have the `.m` extension. Files that declare class and category interfaces or that declare protocols have the `.h` extension typical of header files.

Class, category, and protocol names generally begin with an uppercase letter; the names of methods and instance variables typically begin with a lowercase letter. The names of variables that hold instances usually also begin with lowercase letters.

In Objective-C, identical names that serve different purposes don't clash. Within a class, names can be freely assigned:

- A class can declare methods with the same names as methods in other classes.
- A class can declare instance variables with the same names as variables in other classes.
- An instance method can have the same name as a class method.
- A method can have the same name as an instance variable.
- Method names beginning with "_", a single underscore character, are reserved for use by Apple.

Likewise, protocols and categories of the same class have protected name spaces:

- A protocol can have the same name as a class, a category, or anything else.
- A category of one class can have the same name as a category of another class.

However, class names are in the same name space as global variables and defined types. A program can't have a global variable with the same name as a class.

Document Revision History

This table describes the changes to *The Objective-C Programming Language*.

Date	Notes
2009-10-19	Added discussion of associative references.
2009-08-12	Corrected minor errors.
2009-05-06	Updated article on Mixing Objective-C and C++.
2009-02-04	Updated description of categories.
2008-11-19	Significant reorganization, with several sections moved to a new Runtime Guide.
2008-10-15	Corrected typographical errors.
2008-07-08	Corrected typographical errors.
2008-06-09	Made several minor bug fixes and clarifications, particularly in the "Properties" chapter.
2008-02-08	Extended the discussion of properties to include mutable objects.
2007-12-11	Corrected minor errors.
2007-10-31	Provided an example of fast enumeration for dictionaries and enhanced the description of properties.
2007-07-22	Added references to documents describing new features in Objective-C 2.
2007-03-26	Corrected minor typographical errors.
2007-02-08	Clarified the discussion of sending messages to nil.
2006-12-05	Clarified the description of Code Listing 3-3.
2006-05-23	Moved the discussion of memory management to "Memory Management Programming Guide for Cocoa."
2006-04-04	Corrected minor typographical errors.
2006-02-07	Corrected minor typographical errors.
2006-01-10	Clarified use of the static specifier for global variables used by a class.
2005-10-04	Clarified effect of sending messages to nil; noted use of ".mm" extension to signal Objective-C++ to compiler.

Date	Notes
2005-04-08	Corrected typo in language grammar specification and modified a code example.
	Corrected the grammar for the protocol-declaration-list declaration in "External Declarations".
	Clarified example in Listing 14-1 (page 111).
2004-08-31	Removed function and data structure reference. Added exception and synchronization grammar. Made technical corrections and minor editorial changes.
	Moved function and data structure reference to *Objective-C Runtime Reference*.
	Added examples of thread synchronization approaches to "Synchronizing Thread Execution".
	Clarified when the `initialize` method is called and provided a template for its implementation in "Initializing a Class Object".
	Added exception and synchronization grammar to "Grammar".
	Replaced `conformsTo:` with `conformsToProtocol:` throughout document.
2004-02-02	Corrected typos in "An exception handler".
2003-09-16	Corrected definition of `id`.
2003-08-14	Documented the Objective-C exception and synchronization support available in Mac OS X version 10.3 and later in "Exception Handling and Thread Synchronization".
	Documented the language support for concatenating constant strings in "Compiler Directives" (page 118).
	Moved "Memory Management" before "Retaining Objects".
	Corrected the descriptions for the `Ivar` structure and the `objc_ivar_list` structure.
	Changed the font of *function result* in `class_getInstanceMethod` and `class_getClassMethod`.
	Corrected definition of the term *conform* in the glossary.
	Corrected definition of `method_getArgumentInfo`.
	Renamed from *Inside Mac OS X: The Objective-C Programming Language* to *The Objective-C Programming Language*.
2003-01-01	Documented the language support for declaring constant strings. Fixed several typographical errors. Added an index.

Date	Notes
2002-05-01	Mac OS X 10.1 introduces a compiler for Objective-C++, which allows C++ constructs to be called from Objective-C classes, and vice versa.
	Added runtime library reference material.
	Fixed a bug in the Objective-C language grammar's description of instance variable declarations.
	Updated grammar and section names throughout the book to reduce ambiguities, passive voice, and archaic tone. Restructured some sections to improve cohesiveness.
	Renamed from *Object Oriented Programming and the Objective-C Language* to *Inside Mac OS X: The Objective-C Programming Language*.

Document Revision History

Glossary

abstract class A class that's defined solely so that other classes can inherit from it. Programs don't use instances of an abstract class, only of its subclasses.

abstract superclass Same as abstract class.

adopt In the Objective-C language, a class is said to adopt a protocol if it declares that it implements all the methods in the protocol. Protocols are adopted by listing their names between angle brackets in a class or category declaration.

anonymous object An object of unknown class. The interface to an anonymous object is published through a protocol declaration.

Application Kit A Cocoa framework that implements an application's user interface. The Application Kit provides a basic program structure for applications that draw on the screen and respond to events.

archiving The process of preserving a data structure, especially an object, for later use. An archived data structure is usually stored in a file, but it can also be written to memory, copied to the pasteboard, or sent to another application. In Cocoa, archiving involves writing data to an `NSData` object.

asynchronous message A remote message that returns immediately, without waiting for the application that receives the message to respond. The sending application and the receiving application act independently, and are therefore not "in sync." See also synchronous message.

category In the Objective-C language, a set of method definitions that is segregated from the rest of the class definition. Categories can be used to split a class definition into parts or to add methods to an existing class.

class In the Objective-C language, a prototype for a particular kind of object. A class definition declares instance variables and defines methods for all members of the class. Objects that have the same types of instance variables and have access to the same methods belong to the same class. See also class object.

class method In the Objective-C language, a method that can operate on class objects rather than instances of the class.

class object In the Objective-C language, an object that represents a class and knows how to create new instances of the class. Class objects are created by the compiler, lack instance variables, and can't be statically typed, but otherwise behave like all other objects. As the receiver in a message expression, a class object is represented by the class name.

Cocoa An advanced object-oriented development platform on Mac OS X. Cocoa is a set of frameworks with its primary programming interfaces in Objective-C.

compile time The time when source code is compiled. Decisions made at compile time are constrained by the amount and kind of information encoded in source files.

conform In the Objective-C language, a class is said to conform to a protocol if it (or a superclass) implements the methods declared in the protocol. An instance conforms to a protocol if its class does. Thus, an instance that conforms to a protocol can perform any of the instance methods declared in the protocol.

content view In the Application Kit, the `NSView` object that's associated with the content area of a window—all the area in the window excluding the title bar and border. All other views in the window are arranged in a hierarchy beneath the content view.

delegate An object that acts on behalf of another object.

designated initializer The `init...` method that has primary responsibility for initializing new instances of a class. Each class defines or inherits its own designated initializer. Through messages to `self`, other `init...` methods in the same class directly or indirectly invoke the designated initializer, and the designated initializer, through a message to `super`, invokes the designated initializer of its superclass.

dispatch table Objective-C runtime table that contains entries that associate method selectors with the class-specific addresses of the methods they identify.

distributed objects Architecture that facilitates communication between objects in different address spaces.

dynamic allocation Technique used in C-based languages where the operating system provides memory to a running application as it needs it, instead of when it launches.

dynamic binding Binding a method to a message—that is, finding the method implementation to invoke in response to the message—at runtime, rather than at compile time.

dynamic typing Discovering the class of an object at runtime rather than at compile time.

encapsulation Programming technique that hides the implementation of an operation from its users behind an abstract interface. This allows the implementation to be updated or changed without impacting the users of the interface.

event The direct or indirect report of external activity, especially user activity on the keyboard and mouse.

factory Same as class object.

factory method Same as class method.

factory object Same as class object.

formal protocol In the Objective-C language, a protocol that's declared with the `@protocol` directive. Classes can adopt formal protocols, objects can respond at runtime when asked if they conform to a formal protocol, and instances can be typed by the formal protocols they conform to.

framework A way to package a logically-related set of classes, protocols and functions together with localized strings, on-line documentation, and other pertinent files. Cocoa provides the Foundation framework and the Application Kit framework, among others. Frameworks are sometimes referred to as "kits."

gdb The standard Mac OS X debugging tool.

id In the Objective-C language, the general type for any kind of object regardless of class. `id` is defined as a pointer to an object data structure. It can be used for both class objects and instances of a class.

implementation Part of an Objective-C class specification that defines its implementation. This section defines both public methods as well as private methods—methods that are not declared in the class's interface.

informal protocol In the Objective-C language, a protocol declared as a category, usually as a category of the `NSObject` class. The language gives explicit support to formal protocols, but not to informal ones.

inheritance In object-oriented programming, the ability of a superclass to pass its characteristics (methods and instance variables) on to its subclasses.

inheritance hierarchy In object-oriented programming, the hierarchy of classes that's defined by the arrangement of superclasses and subclasses. Every class (except root classes such as `NSObject`) has a superclass, and any class may have an unlimited number of subclasses. Through its superclass, each class inherits from those above it in the hierarchy.

instance In the Objective-C language, an object that belongs to (is a member of) a particular class. Instances are created at runtime according to the specification in the class definition.

instance method In the Objective-C language, any method that can be used by an instance of a class rather than by the class object.

instance variable In the Objective-C language, any variable that's part of the internal data structure of an instance. Instance variables are declared in a class definition and become part of all objects that are members of or inherit from the class.

interface Part of an Objective-C class specification that declares its public interface, which include its superclass name, instances variables, and public-method prototypes.

Interface Builder A tool that lets you graphically specify your application's user interface. It sets up the corresponding objects for you and makes it easy for you to establish connections between these objects and your own code where needed.

introspection The ability of an object to reveal information about itself as an object—such as its class and superclass, the messages it can respond to, and the protocols it conforms to.

key window The window in the active application that receives keyboard events and is the focus of user activity.

link time The time when files compiled from different source modules are linked into a single program. Decisions made by the linker are constrained by the compiled code and ultimately by the information contained in source code.

localize To adapt an application to work under various local conditions—especially to have it use a language selected by the user. Localization entails freeing application code from language-specific and culture-specific references and making it able to import localized resources (such as character strings, images, and sounds). For example, an application localized in Spanish would display "Salir" in the application menu. In Italian, it would be "Esci," in German "Verlassen," and in English "Quit."

main event loop The principal control loop for applications that are driven by events. From the time it's launched until the moment it's terminated, an application gets one keyboard or mouse event after another from the Window Manager and responds to them, waiting between events if the next event isn't ready. In the Application Kit, the `NSApplication` object runs the main event loop.

menu A small window that displays a list of commands. Only menus for the active application are visible on-screen.

message In object-oriented programming, the method selector (name) and accompanying arguments that tell the receiving object in a message expression what to do.

message expression In object-oriented programming, an expression that sends a message to an object. In the Objective-C language, message expressions are enclosed within square brackets and consist of a receiver followed by a message (method selector and arguments).

method In object-oriented programming, a procedure that can be executed by an object.

multiple inheritance In object-oriented programming, the ability of a class to have more than one superclass—to inherit from different sources and thus combine separately-defined behaviors in a single class. Objective-C doesn't support multiple inheritance.

mutex Also known as mutual exclusion semaphore. Used to synchronize thread execution.

name space A logical subdivision of a program within which all names must be unique. Symbols in one name space won't conflict with identically named symbols in another name space. For example, in Objective-C, the instance methods of each class are in a separate name space, as are the class methods and instance variables.

nil In the Objective-C language, an object `id` with a value of 0.

object A programming unit that groups together a data structure (instance variables) and the operations (methods) that can use or affect that data. Objects are the principal building blocks of object-oriented programs.

outlet An instance variable that points to another object. Outlet instance variables are a way for an object to keep track of the other objects to which it may need to send messages.

polymorphism In object-oriented programming, the ability of different objects to respond, each in its own way, to the same message.

procedural programming language A language, like C, that organizes a program as a set of procedures that have definite beginnings and ends.

protocol In the Objective-C language, the declaration of a group of methods not associated with any particular class. See also formal protocol and informal protocol.

receiver In object-oriented programming, the object that is sent a message.

reference counting Memory-management technique in which each entity that claims ownership of an object increments the object's reference count and later decrements it. When the object's reference count reaches zero, the object is deallocated. This technique allows one instance of an object to be safely shared among several other objects.

remote message A message sent from one application to an object in another application.

remote object An object in another application, one that's a potential receiver for a remote message.

runtime The time after a program is launched and while it's running. Decisions made at runtime can be influenced by choices the user makes.

selector In the Objective-C language, the name of a method when it's used in a source-code message to an object, or the unique identifier that replaces the name when the source code is compiled. Compiled selectors are of type `SEL`.

static typing In the Objective-C language, giving the compiler information about what kind of object an instance is, by typing it as a pointer to a class.

subclass In the Objective-C language, any class that's one step below another class in the inheritance hierarchy. Occasionally used more generally to mean any class that inherits from another class, and sometimes also used as a verb to mean the process of defining a subclass of another class.

superclass In the Objective-C language, a class that's one step above another class in the inheritance hierarchy; the class through which a subclass inherits methods and instance variables.

surrogate An object that stands in for and forwards messages to another object.

synchronous message A remote message that doesn't return until the receiving application finishes responding to the message. Because the application that sends the message waits for an acknowledgment or return information from the receiving application, the two applications are kept "in sync." See also asynchronous message.

Index

Symbols

A

B

C

D

E

F

G

H

I

M

N

O

P

R

S

T

U

V

Y