

An Introduction to Object-Oriented Design

Lane W. Surface

4 March 2019

Among many of the frameworks, APIs, languages, and other skill sets you will be expected to know as a professional programmer, a sound knowledge of the principles of object-oriented design (henceforth referred to as OOD) will be invaluable in your career. Though it may seem simple on the surface, even professionals struggle with the decisions that come with a well-designed and easy to maintain system which strives to follow the best OOD practices. This document will serve primarily as a brief introduction for developers with no or limited experience in this domain.

1 Classes, Methods, and State: It All Ties Into Encapsulation

A recurring theme which you will come to see in the evolution of programming languages, and thus software development in general, is the tendency for a language to make the solution to a given problem both easier to write and easier to read. Assembly languages were originally the de-facto standard for writing applications, and many were written solely in it, including important and critical software like that which ran on the Apollo Guidance Computers and put men on the moon. From these struggles, a new way of writing software was proposed, and higher-level languages based on the notion of functions were created. From these languages has come one of the newest developments in the programming language world, which is based around the notion of the object.

But what is an object anyway? For how often the term is thrown around, there really isn't much of an attempt to explain it or the principles on which it is based. An object is simply a component of an object-oriented system; this means that there are many objects which make up an object-based software system. Each of these objects interacts with the others which make up this application to create a working program.

Let's expound this definition further: Each object itself has both state and behavior. This state, often called fields of the object, resides in class-level variables which are instance-dependent. In some languages these are called instance variables. State is a property of the object itself. Say you have a human object. This object might have state such as hair color, race, ethnicity, nationality, etc.

Although the state of an object is important, it is mostly internal (and should be kept hidden; we'll address this further in a moment). The primary way that objects should expose their functionality to other objects in a system is through instance-dependent functions, which are often referred to as methods of that object. These methods use the state of their object to determine the action that should be taken, without requiring the objects which call that function to know about the state itself. Therefore, calling a method such as `pray()` on a

human object should determine the way that it prays by these factors, without a caller needing to know all of the internal details.

Now I've skipped over an important distinction to make, as I wanted to avoid confusing you up front. How do we create and address the objects in our system? This state is not intrinsic, and we have to define it somewhere in our application. We do so by defining classes. Classes are a template for the construction of an object. In other words, the class defines the how, and when we construct an object from a class, we define the what. A class will determine the kinds of things which we need to know about. Taking from the earlier example of a human object, the human class will tell the compiler that we need each of our human objects to have an integer for their age, string for their name, etc. Then when we initialize a human object (as there may be many in our system), we provide the values of these fields. It is only after an object has been constructed and initialized that we can call the methods which it defines. The type of the object is the same as the class which defines it, and the name that the variable is given when constructed is the object itself.

You may be wondering why it's considered good practice to not expose the state of an object. The reasons are twofold. First, forcing a client to interact with the state of an object increases the complexity of your code. The purpose of designing an object-oriented system is to hide away information that is not important to the rest of the system. If the only thing we need to know is what behavior an object supports (i.e., the methods which it defines in its public interface), then we are much less likely to be surprised by the code that we write. Another reason is because, as software is maintained over its lifetime, the fields and their respective types are very likely to change; it is not often that we write code once and are done with it. The majority of the lifetime of a program is in maintaining it. We will see that it is much easier to make modifications to the state and behavior of an object when we do not expose this state directly.