BJARNE STROUSTRUP

THE CREATOR OF C++

Using
C++11
and
C++14

PROGRAMMING

Principles and Practice Using C++

SECOND EDITION

About This eBook

ePUB is an open, industry-standard format for eBooks. However, support of ePUB and its many features varies across reading devices and applications. Use your device or app settings to customize the presentation to your liking. Settings that you can customize often include font, font size, single or double column, landscape or portrait mode, and figures that you can click or tap to enlarge. For additional information about the settings and features on your reading device or app, visit the device manufacturer's Web site.

Many titles include programming code or configuration examples. To optimize the presentation of these elements, view the eBook in single-column, landscape mode and adjust the font size to the smallest setting. In addition to presenting code and configurations in the reflowable text format, we have included images of the code that mimic the presentation found in the print book; therefore, where the reflowable format may compromise the presentation of the code listing, you will see a "Click here to view code image" link. Click the link to view the print-fidelity code image. To return to the previous page viewed, click the Back button on your device or app.

Programming: Principles and Practice Using C++

Second Edition

Bjarne Stroustrup

Addison-Wesley
Upper Saddle River, NJ • Boston • Indianapolis • San Francisco
New York • Toronto • Montreal • London • Munich • Paris • Madrid
Capetown • Sydney • Tokyo • Singapore • Mexico City

A complete list of photo sources and credits appears on pages 1273–1274.

Library of Congress Cataloging-in-Publication Data

Stroustrup, Bjarne, author.
Programming : principles and practice using C++ / Bjarne Stroustrup. — Second edition.
pages cm
Includes bibliographical references and index.
ISBN 978-0-321-99278-9 (pbk. : alk. paper)
1. C++ (Computer program language) I. Title.
QA76.73.C153S82 2014
005.13'3—dc23

2014004197

ISBN-13: 978-0-321-99278-9
ISBN-10: 0-321-99278-4
Text printed in the United States on recycled paper at RR Donnelley in Crawfordsville, Indiana.
First printing, May 2014

Contents

Preface

"Damn the torpedoes! Full speed ahead."

—Admiral Farragut

Programming is the art of expressing solutions to problems so that a computer can execute those solutions. Much of the effort in programming is spent finding and refining solutions. Often, a problem is only fully understood through the process of programming a solution for it.

This book is for someone who has never programmed before but is willing to work hard to learn. It helps you understand the principles and acquire the practical skills of programming using the C++ programming language. My aim is for you to gain sufficient knowledge and experience to perform simple useful programming tasks using the best up-to-date techniques. How long will that take? As part of a first-year university course, you can work through this book in a semester (assuming that you have a workload of four courses of average difficulty). If you work by yourself, don't expect to spend less time than that (maybe 15 hours a week for 14 weeks).

Three months may seem a long time, but there's a lot to learn and you'll be writing your first simple programs after about an hour. Also, all learning is gradual: each chapter introduces new useful concepts and illustrates them with examples inspired by real-world uses. Your ability to express ideas in code — getting a computer to do what you want it to do — gradually and steadily increases as you go along. I never say, "Learn a month's worth of theory and then see if you can use it."

Why would you want to program? Our civilization runs on software. Without understanding software you are reduced to believing in "magic" and will be locked out of many of the most interesting, profitable, and socially useful technical fields of work. When I talk about programming, I think of the whole spectrum of computer programs from personal computer applications with GUIs (graphical user interfaces), through engineering calculations and embedded systems control applications (such as digital cameras, cars, and cell phones), to text manipulation applications as found in many humanities and business applications. Like mathematics, programming — when done well — is a valuable intellectual exercise that sharpens our ability to think. However, thanks to feedback from the computer, programming is more concrete than most forms of math, and therefore accessible to more people. It is a way to reach out and change the world — ideally for the better. Finally, programming can be great fun.

Why C++? You can't learn to program without a programming language, and C++ directly supports the key concepts and techniques used in real-world software. C++ is one of the most widely used programming languages, found in an unsurpassed range of application areas. You find C++ applications everywhere from the bottom of the oceans to the surface of Mars. C++ is precisely and comprehensively defined by a nonproprietary international standard. Quality and/or free implementations are available on every kind of computer. Most of the programming concepts that you will learn using C++ can be used directly in other languages, such as C, C#, Fortran, and Java. Finally, I simply like C++ as a language for writing elegant and efficient code.

This is not the easiest book on beginning programming; it is not meant to be. I just aim for it to be the easiest book from which you can learn the basics of real-world programming. That's quite an ambitious goal because much modern software relies on techniques considered advanced just a few years ago.

My fundamental assumption is that you want to write programs for the use of others, and to do so responsibly, providing a decent level of system quality; that is, I assume that you want to achieve a level of professionalism. Consequently, I chose the topics for this book to cover what is needed to get started with real-world programming, not just what is easy to teach and learn. If you need a technique to get basic work done right, I describe it, demonstrate concepts and language facilities needed to support the technique, provide exercises for it, and expect you to work on those exercises. If you just want to understand toy programs, you can get along with far less than I present. On the other hand, I won't waste your time with material of marginal practical importance. If an idea is explained here, it's because you'll almost certainly need it.

If your desire is to use the work of others without understanding how things are done and without adding significantly to the code yourself, this book is not for you. If so, please consider whether you would be better served by another book and another language. If that is approximately your view of programming, please also consider from where you got that view and whether it in fact is adequate for your needs. People often underestimate the complexity of programming as well as its value. I would hate for you to acquire a dislike for programming because of a mismatch between what you need and the part of the software reality I describe. There are many parts of the "information technology" world that do not require knowledge of programming. This book is aimed to serve those who do want to write or understand nontrivial programs.

Because of its structure and practical aims, this book can also be used as a second book on programming for someone who already knows a bit of C++ or for someone who programs in another language and wants to learn C++. If you fit into one of

those categories, I refrain from guessing how long it will take you to read this book, but I do encourage you to do many of the exercises. This will help you to counteract the common problem of writing programs in older, familiar styles rather than adopting newer techniques where these are more appropriate. If you have learned C++ in one of the more traditional ways, you'll find something surprising and useful before you reach Chapter 7. Unless your name is Stroustrup, what I discuss here is not "your father's C++."

Programming is learned by writing programs. In this, programming is similar to other endeavors with a practical component. You cannot learn to swim, to play a musical instrument, or to drive a car just from reading a book — you must practice. Nor can you learn to program without reading and writing lots of code. This book focuses on code examples closely tied to explanatory text and diagrams. You need those to understand the ideals, concepts, and principles of programming and to master the language constructs used to express them. That's essential, but by itself, it will not give you the practical skills of programming. For that, you need to do the exercises and get used to the tools for writing, compiling, and running programs. You need to make your own mistakes and learn to correct them. There is no substitute for writing code. Besides, that's where the fun is!

On the other hand, there is more to programming — much more — than following a few rules and reading the manual. This book is emphatically not focused on "the syntax of C++." Understanding the fundamental ideals, principles, and techniques is the essence of a good programmer. Only well-designed code has a chance of becoming part of a correct, reliable, and maintainable system. Also, "the fundamentals" are what last: they will still be essential after today's languages and tools have evolved or been replaced.

What about computer science, software engineering, information technology, etc.? Is that all programming? Of course not! Programming is one of the fundamental topics that underlie everything in computer-related fields, and it has a natural place in a balanced course of computer science. I provide brief introductions to key concepts and techniques of algorithms, data structures, user interfaces, data processing, and software engineering. However, this book is not a substitute for a thorough and balanced study of those topics.

Code can be beautiful as well as useful. This book is written to help you see that, to understand what it means for code to be beautiful, and to help you to master the principles and acquire the practical skills to create such code. Good luck with programming!

A note to students

Of the many thousands of first-year students we have taught so far using this book at Texas A&M University, about 60% had programmed before and about 40% had never seen a line of code in their lives. Most succeeded, so you can do it, too.

You don't have to read this book as part of a course. The book is widely used for self-study. However, whether you work your way through as part of a course or independently, try to work with others. Programming has an — unfair — reputation as a lonely activity. Most people work better and learn faster when they are part of a group with a common aim. Learning together and discussing problems with friends is not cheating! It is the most efficient — as well as most pleasant — way of making progress. If nothing else, working with friends forces you to articulate your ideas, which is just about the most efficient way of testing your understanding and making sure you remember. You don't actually have to personally discover the answer to every obscure language and programming environment problem. However, please don't cheat yourself by not doing the drills and a fair number of exercises (even if no teacher forces you to do them). Remember: programming is (among other things) a practical skill that you need to practice to master. If you don't write code (do several exercises for each chapter), reading this book will be a pointless theoretical exercise.

Most students — especially thoughtful good students — face times when they wonder whether their hard work is worthwhile. When (not if) this happens to you, take a break, reread this Preface, and look at Chapter 1 ("Computers, People, and Programming") and Chapter 22 ("Ideals and History"). There, I try to articulate what I find exciting about programming and why I consider it a crucial tool for making a positive contribution to the world. If you wonder about my teaching philosophy and general approach, have a look at Chapter 0 ("Notes to the Reader").

You might find the weight of this book worrying, but it should reassure you that part of the reason for the heft is that I prefer to repeat an explanation or add an example rather than have you search for the one and only explanation. The other major reason is that the second half of the book is reference material and "additional material" presented for you to explore only if you are interested in more information about a specific area of programming, such as embedded systems programming, text analysis, or numerical computation.

And please don't be too impatient. Learning any major new and valuable skill takes time and is worth it.

A note to teachers

No. This is not a traditional Computer Science 101 course. It is a book about how to construct working software. As such, it

leaves out much of what a computer science student is traditionally exposed to (Turing completeness, state machines, discrete math, Chomsky grammars, etc.). Even hardware is ignored on the assumption that students have used computers in various ways since kindergarten. This book does not even try to mention most important CS topics. It is about programming (or more generally about how to develop software), and as such it goes into more detail about fewer topics than many traditional courses. It tries to do just one thing well, and computer science is not a one-course topic. If this book/course is used as part of a computer science, computer engineering, electrical engineering (many of our first students were EE majors), information science, or whatever program, I expect it to be taught alongside other courses as part of a well-rounded introduction.

Please read Chapter 0 ("Notes to the Reader") for an explanation of my teaching philosophy, general approach, etc. Please try to convey those ideas to your students along the way.

ISO standard C++

C++ is defined by an ISO standard. The first ISO C++ standard was ratified in 1998, so that version of C++ is known as C++98. I wrote the first edition of this book while working on the design of C++11. It was most frustrating not to be able to use the novel features (such as uniform initialization, range-**for**-loops, move semantics, lambdas, and concepts) to simplify the presentation of principles and techniques. However, the book was designed with C++11 in mind, so it was relatively easy to "drop in" the features in the contexts where they belonged. As of this writing, the current standard is C++11 from 2011, and facilities from the upcoming 2014 ISO standard, C++14, are finding their way into mainstream C++ implementations. The language used in this book is C++11 with a few C++14 features. For example, if your compiler complains about

Click here to view code image

```
vector<int> v1;
vector<int> v2 {v1};      // C++14-style copy construction
```

use

Click here to view code image

```
vector<int> v1;
vector<int> v2 = v1;      // C++98-style copy construction
```

instead.

If your compiler does not support C++11, get a new compiler. Good, modern C++ compilers can be downloaded from a variety of suppliers; see www.stroustrup.com/compilers.html. Learning to program using an earlier and less supportive version of the language can be unnecessarily hard.

Support

The book's support website, www.stroustrup.com/Programming, contains a variety of material supporting the teaching and learning of programming using this book. The material is likely to be improved with time, but for starters, you can find

- Slides for lectures based on the book
- An instructor's guide
- Header files and implementations of libraries used in the book
- Code for examples in the book
- Solutions to selected exercises
- Potentially useful links
- Errata

Suggestions for improvements are always welcome.

Acknowledgments

I'd especially like to thank my late colleague and co-teacher Lawrence "Pete" Petersen for encouraging me to tackle the task of teaching beginners long before I'd otherwise have felt comfortable doing that, and for supplying the practical teaching experience to make the course succeed. Without him, the first version of the course would have been a failure. We worked together on the first versions of the course for which this book was designed and together taught it repeatedly, learning from our experiences, improving the course and the book. My use of "we" in this book initially meant "Pete and me."

Thanks to the students, teaching assistants, and peer teachers of ENGR 112, ENGR 113, and CSCE 121 at Texas A&M University who directly and indirectly helped us construct this book, and to Walter Daugherity, Hyunyoung Lee, Teresa Leyk, Ronnie Ward, and Jennifer Welch, who have also taught the course. Also thanks to Damian Dechev, Tracy Hammond, Arne

Tolstrup Madsen, Gabriel Dos Reis, Nicholas Stroustrup, J. C. van Winkel, Greg Versoonder, Ronnie Ward, and Leor Zolman for constructive comments on drafts of this book. Thanks to Mogens Hansen for explaining about engine control software. Thanks to Al Aho, Stephen Edwards, Brian Kernighan, and Daisy Nguyen for helping me hide away from distractions to get writing done during the summers.

Thanks to Art Werschulz for many constructive comments based on his use of the first edition of this book in courses at Fordham University in New York City and to Nick Maclaren for many detailed comments on the exercises based on his use of the first edition of this book at Cambridge University. His students had dramatically different backgrounds and professional needs from the TAMU first-year students.

Thanks to the reviewers that Addison-Wesley found for me. Their comments, mostly based on teaching either C++ or Computer Science 101 at the college level, have been invaluable: Richard Enbody, David Gustafson, Ron McCarty, and K. Narayanaswamy. Also thanks to my editor, Peter Gordon, for many useful comments and (not least) for his patience. I'm very grateful to the production team assembled by Addison-Wesley; they added much to the quality of the book: Linda Begley (proofreader), Kim Arney (compositor), Rob Mauhar (illustrator), Julie Nahil (production editor), and Barbara Wood (copy editor).

Thanks to the translators of the first edition, who found many problems and helped clarify many points. In particular, Loïc Joly and Michel Michaud did a thorough technical review of the French translation that led to many improvements.

I would also like to thank Brian Kernighan and Doug McIlroy for setting a very high standard for writing about programming, and Dennis Ritchie and Kristen Nygaard for providing valuable lessons in practical language design.

0. Notes to the Reader

"When the terrain disagrees with the map, trust the terrain."

—Swiss army proverb

This chapter is a grab bag of information; it aims to give you an idea of what to expect from the rest of the book. Please skim through it and read what you find interesting. A teacher will find most parts immediately useful. If you are reading this book without the benefit of a good teacher, please don't try to read and understand everything in this chapter; just look at "The structure of this book" and the first part of the "A philosophy of teaching and learning" sections. You may want to return and reread this chapter once you feel comfortable writing and executing small programs.

0.1 The structure of this book

This book consists of four parts and a collection of appendices:

- *Part I*, *"The Basics,"* presents the fundamental concepts and techniques of programming together with the C++ language and library facilities needed to get started writing code. This includes the type system, arithmetic operations, control structures, error handling, and the design, implementation, and use of functions and user-defined types.
- *Part II*, *"Input and Output,"* describes how to get numeric and text data from the keyboard and from files, and how to produce corresponding output to the screen and to files. Then, it shows how to present numeric data, text, and geometric shapes as graphical output, and how to get input into a program from a graphical user interface (GUI).
- *Part III*, *"Data and Algorithms,"* focuses on the C++ standard library's containers and algorithms framework (the STL, standard template library). It shows how containers (such as **vector**, **list**, and **map**) are implemented (using pointers, arrays, dynamic memory, exceptions, and templates) and used. It also demonstrates the design and use of standard library algorithms (such as **sort**, **find**, and **inner_product**).
- *Part IV*, *"Broadening the View,"* offers a perspective on programming through a discussion of ideals and history, through examples (such as matrix computation, text manipulation, testing, and embedded systems programming), and through a brief description of the C language.
- *Appendices* provide useful information that doesn't fit into a tutorial presentation, such as surveys of C++ language and standard library facilities, and descriptions of how to get started with an integrated development environment (IDE) and a graphical user interface (GUI) library.

Unfortunately, the world of programming doesn't really fall into four cleanly separated parts. Therefore, the "parts" of this book provide only a coarse classification of topics. We consider it a useful classification (obviously, or we wouldn't have used it), but reality has a way of escaping neat classifications. For example, we need to use input operations far sooner than we can give a thorough explanation of C++ standard I/O streams (input/output streams). Where the set of topics needed to present an idea conflicts with the overall classification, we explain the minimum needed for a good presentation, rather than just

referring to the complete explanation elsewhere. Rigid classifications work much better for manuals than for tutorials.

The order of topics is determined by programming techniques, rather than programming language features; see §0.2. For a presentation organized around language features, see Appendix A.

To ease review and to help you if you miss a key point during a first reading where you have yet to discover which kind of information is crucial, we place three kinds of "alert markers" in the margin:

- Blue: concepts and techniques (this paragraph is an example of that)
- Green: advice
- Red: warning

0.1.1 General approach

In this book, we address you directly. That is simpler and clearer than the conventional "professional" indirect form of address, as found in most scientific papers. By "you" we mean "you, the reader," and by "we" we refer either to "ourselves, the author and teachers," or to you and us working together through a problem, as we might have done had we been in the same room.

This book is designed to be read chapter by chapter from the beginning to the end. Often, you'll want to go back to look at something a second or a third time. In fact, that's the only sensible approach, as you'll always dash past some details that you don't yet see the point in. In such cases, you'll eventually go back again. However, despite the index and the cross-references, this is not a book that you can open to any page and start reading with any expectation of success. Each section and each chapter assume understanding of what came before.

Each chapter is a reasonably self-contained unit, meant to be read in "one sitting" (logically, if not always feasible on a student's tight schedule). That's one major criterion for separating the text into chapters. Other criteria include that a chapter is a suitable unit for drills and exercises and that each chapter presents some specific concept, idea, or technique. This plurality of criteria has left a few chapters uncomfortably long, so please don't take "in one sitting" too literally. In particular, once you have thought about the review questions, done the drill, and worked on a few exercises, you'll often find that you have to go back to reread a few sections and that several days have gone by. We have clustered the chapters into "parts" focused on a major topic, such as input/output. These parts make good units of review.

Common praise for a textbook is "It answered all my questions just as I thought of them!" That's an ideal for minor technical questions, and early readers have observed the phenomenon with this book. However, that cannot be the whole ideal. We raise questions that a novice would probably not think of. We aim to ask and answer questions that you need to consider when writing quality software for the use of others. Learning to ask the right (often hard) questions is an essential part of learning to think as a programmer. Asking only the easy and obvious questions would make you feel good, but it wouldn't help make you a programmer.

We try to respect your intelligence and to be considerate about your time. In our presentation, we aim for professionalism rather than cuteness, and we'd rather understate a point than hype it. We try not to exaggerate the importance of a programming technique or a language feature, but please don't underestimate a simple statement like "This is often useful." If we quietly emphasize that something is important, we mean that you'll sooner or later waste days if you don't master it. Our use of humor is more limited than we would have preferred, but experience shows that people's ideas of what is funny differ dramatically and that a failed attempt at humor can be confusing.

We do not pretend that our ideas or the tools offered are perfect. No tool, library, language, or technique is "the solution" to all of the many challenges facing a programmer. At best, it can help you to develop and express your solution. We try hard to avoid "white lies"; that is, we refrain from oversimplified explanations that are clear and easy to understand, but not true in the context of real languages and real problems. On the other hand, this book is not a reference; for more precise and complete descriptions of C++, see Bjarne Stroustrup, *The C++ Programming Language, Fourth Edition* (Addison-Wesley, 2013), and the ISO C++ standard.

0.1.2 Drills, exercises, etc.

Programming is not just an intellectual activity, so writing programs is necessary to master programming skills. We provide two levels of programming practice:

- *Drills:* A drill is a very simple exercise devised to develop practical, almost mechanical skills. A drill usually consists of a sequence of modifications of a single program. You should do every drill. A drill is not asking for deep understanding, cleverness, or initiative. We consider the drills part of the basic fabric of the book. If you haven't done the drills, you have not "done" the book.
- *Exercises:* Some exercises are trivial and others are very hard, but most are intended to leave some scope for initiative and imagination. If you are serious, you'll do quite a few exercises. At least do enough to know which are difficult for you. Then do a few more of those. That's how you'll learn the most. The exercises are meant to be manageable without exceptional cleverness, rather than to be tricky puzzles. However, we hope that we have provided exercises that are hard enough to challenge anybody and enough exercises to exhaust even the best student's available time. We do not expect you to do them all, but feel free to try.

In addition, we recommend that you (every student) take part in a small project (and more if time allows for it). A project is intended to produce a complete useful program. Ideally, a project is done by a small group of people (e.g., three people) working together for about a month while working through the chapters in Part III. Most people find the projects the most fun and what ties everything together.

Some people like to put the book aside and try some examples before reading to the end of a chapter; others prefer to read ahead to the end before trying to get code to run. To support readers with the former preference, we provide simple suggestions for practical work labeled "**Try this**" at natural breaks in the text. A **Try this** is generally in the nature of a drill focused narrowly on the topic that precedes it. If you pass a **Try this** without trying — maybe because you are not near a computer or you find the text riveting — do return to it when you do the chapter drill; a **Try this** either complements the chapter drill or is a part of it.

At the end of each chapter you'll find a set of review questions. They are intended to point you to the key ideas explained in the chapter. One way to look at the review questions is as a complement to the exercises: the exercises focus on the practical aspects of programming, whereas the review questions try to help you articulate the ideas and concepts. In that, they resemble good interview questions.

The "Terms" section at the end of each chapter presents the basic vocabulary of programming and of C++. If you want to understand what people say about programming topics and to articulate your own ideas, you should know what each means.

Learning involves repetition. Our ideal is to make every important point at least twice and to reinforce it with exercises.

0.1.3 What comes after this book?

At the end of this book, will you be an expert at programming and at C++? Of course not! When done well, programming is a subtle, deep, and highly skilled art building on a variety of technical skills. You should no more expect to be an expert at programming in four months than you should expect to be an expert in biology, in math, in a natural language (such as Chinese, English, or Danish), or at playing the violin in four months — or in half a year, or a year. What you should hope for, and what you can expect if you approach this book seriously, is to have a really good start that allows you to write relatively simple useful programs, to be able to read more complex programs, and to have a good conceptual and practical background for further work.

The best follow-up to this initial course is to work on a real project developing code to be used by someone else. After that, or (even better) in parallel with a real project, read either a professional-level general textbook (such as Stroustrup, *The C++ Programming Language*), a more specialized book relating to the needs of your project (such as Qt for GUI, or ACE for distributed programming), or a textbook focusing on a particular aspect of C++ (such as Koenig and Moo, *Accelerated C++*; Sutter's *Exceptional C++*; or Gamma et al., *Design Patterns*). For more references, see §0.6 or the Bibliography section at the back of the book.

Eventually, you should learn another programming language. We don't consider it possible to be a professional in the realm of software — even if you are not primarily a programmer — without knowing more than one language.

0.2 A philosophy of teaching and learning

What are we trying to help you learn? And how are we approaching the process of teaching? We try to present the minimal concepts, techniques, and tools for you to do effective practical programs, including

- Program organization
- Debugging and testing
- Class design
- Computation
- Function and algorithm design
- Graphics (two-dimensional only)
- Graphical user interfaces (GUIs)
- Text manipulation
- Regular expression matching
- Files and stream input and output (I/O)
- Memory management
- Scientific/numerical/engineering calculations
- Design and programming ideals
- The C++ standard library
- Software development strategies
- C-language programming techniques

Working our way through these topics, we cover the programming techniques called procedural programming (as with the C programming language), data abstraction, object-oriented programming, and generic programming. The main topic of this book is *programming,* that is, the ideals, techniques, and tools of expressing ideas in code. The C++ programming language is our main tool, so we describe many of C++'s facilities in some detail. But please remember that C++ is just a tool, rather than the main topic of this book. This is "programming using C++," not "C++ with a bit of programming theory."

Each topic we address serves at least two purposes: it presents a technique, concept, or principle and also a practical language or library feature. For example, we use the interface to a two-dimensional graphics system to illustrate the use of classes and inheritance. This allows us to be economical with space (and your time) and also to emphasize that programming is more than simply slinging code together to get a result as quickly as possible. The C++ standard library is a major source of such "double duty" examples — many even do triple duty. For example, we introduce the standard library **vector**, use it to illustrate widely useful design techniques, and show many of the programming techniques used to implement it. One of our aims is to show you how major library facilities are implemented and how they map to hardware. We insist that craftsmen must understand their tools, not just consider them "magical."

Some topics will be of greater interest to some programmers than to others. However, we encourage you not to prejudge your needs (how would you know what you'll need in the future?) and at least look at every chapter. If you read this book as part of a course, your teacher will guide your selection.

We characterize our approach as "depth-first." It is also "concrete-first" and "concept-based." First, we quickly (well, relatively quickly, Chapters 1–11) assemble a set of skills needed for writing small practical programs. In doing so, we present a lot of tools and techniques in minimal detail. We focus on simple concrete code examples because people grasp the concrete faster than the abstract. That's simply the way most humans learn. At this initial stage, you should not expect to understand every little detail. In particular, you'll find that trying something slightly different from what just worked can have "mysterious" effects. Do try, though! And please do the drills and exercises we provide. Just remember that early on you just don't have the concepts and skills to accurately estimate what's simple and what's complicated; expect surprises and learn from them.

We move fast in this initial phase — we want to get you to the point where you can write interesting programs as fast as possible. Someone will argue, "We must move slowly and carefully; we must walk before we can run!" But have you ever watched a baby learning to walk? Babies really do run by themselves before they learn the finer skills of slow, controlled walking. Similarly, you will dash ahead, occasionally stumbling, to get a feel of programming before slowing down to gain the necessary finer control and understanding. You must run before you can walk!

It is essential that you don't get stuck in an attempt to learn "everything" about some language detail or technique. For

example, you could memorize all of C++'s built-in types and all the rules for their use. Of course you could, and doing so might make you feel knowledgeable. However, it would not make you a programmer. Skipping details will get you "burned" occasionally for lack of knowledge, but it is the fastest way to gain the perspective needed to write good programs. Note that our approach is essentially the one used by children learning their native language and also the most effective approach used to teach foreign languages. We encourage you to seek help from teachers, friends, colleagues, instructors, Mentors, etc. on the inevitable occasions when you are stuck. Be assured that nothing in these early chapters is fundamentally difficult. However, much will be unfamiliar and might therefore feel difficult at first.

Later, we build on the initial skills to broaden your base of knowledge and skills. We use examples and exercises to solidify your understanding, and to provide a conceptual base for programming.

We place a heavy emphasis on ideals and reasons. You need ideals to guide you when you look for practical solutions — to know when a solution is good and principled. You need to understand the reasons behind those ideals to understand why they should be your ideals, why aiming for them will help you and the users of your code. Nobody should be satisfied with "because that's the way it is" as an explanation. More importantly, an understanding of ideals and reasons allows you to generalize from what you know to new situations and to combine ideas and tools in novel ways to address new problems. Knowing "why" is an essential part of acquiring programming skills. Conversely, just memorizing lots of poorly understood rules and language facilities is limiting, a source of errors, and a massive waste of time. We consider your time precious and try not to waste it.

Many C++ language-technical details are banished to appendices and manuals, where you can look them up when needed. We assume that you have the initiative to search out information when needed. Use the index and the table of contents. Don't forget the online help facilities of your compiler, and the web. Remember, though, to consider every web resource highly suspect until you have reason to believe better of it. Many an authoritative-looking website is put up by a programming novice or someone with something to sell. Others are simply outdated. We provide a collection of links and information on our support website: www.stroustrup.com/Programming.

Please don't be too impatient for "realistic" examples. Our ideal example is the shortest and simplest code that directly illustrates a language facility, a concept, or a technique. Most real-world examples are far messier than ours, yet do not consist of more than a combination of what we demonstrate. Successful commercial programs with hundreds of thousands of lines of code are based on techniques that we illustrate in a dozen 50-line programs. The fastest way to understand real-world code is through a good understanding of the fundamentals.

On the other hand, we do not use "cute examples involving cuddly animals" to illustrate our points. We assume that you aim to write real programs to be used by real people, so every example that is not presented as language-technical is taken from a real-world use. Our basic tone is that of professionals addressing (future) professionals.

0.2.1 The order of topics

There are many ways to teach people how to program. Clearly, we don't subscribe to the popular "the way I learned to program is the best way to learn" theories. To ease learning, we early on present topics that would have been considered advanced only a few years ago. Our ideal is for the topics we present to be driven by problems you meet as you learn to program, to flow smoothly from topic to topic as you increase your understanding and practical skills. The major flow of this book is more like a story than a dictionary or a hierarchical order.

It is impossible to learn all the principles, techniques, and language facilities needed to write a program at once. Consequently, we have to choose a subset of principles, techniques, and features to start with. More generally, a textbook or a course must lead students through a series of subsets. We consider it our responsibility to select topics and to provide emphasis. We can't just present everything, so we must choose; what we leave out is at least as important as what we leave in — at each stage of the journey.

For contrast, it may be useful for you to see a list of (severely abbreviated) characterizations of approaches that we decided not to take:

- *"C first":* This approach to learning C++ is wasteful of students' time and leads to poor programming practices by forcing students to approach problems with fewer facilities, techniques, and libraries than necessary. C++ provides stronger type checking than C, a standard library with better support for novices, and exceptions for error handling.
- *Bottom-up:* This approach distracts from learning good and effective programming practices. By forcing students to solve problems with insufficient support from the language and libraries, it promotes poor and wasteful programming

practices.

- *"If you present something, you must present it fully":* This approach implies a bottom-up approach (by drilling deeper and deeper into every topic touched). It bores novices with technical details they have no interest in and quite likely will not need for years to come. Once you can program, you can look up technical details in a manual. Manuals are good at that, whereas they are awful for initial learning of concepts.
- *Top-down:* This approach, working from first principles toward details, tends to distract readers from the practical aspects of programming and force them to concentrate on high-level concepts before they have any chance of appreciating their importance. For example, you simply can't appreciate proper software development principles before you have learned how easy it is to make a mistake in a program and how hard it can be to correct it.
- *"Abstract first":* Focusing on general principles and protecting the student from nasty real-world constraints can lead to a disdain for real-world problems, languages, tools, and hardware constraints. Often, this approach is supported by "teaching languages" that cannot be used later and (deliberately) insulate students from hardware and system concerns.
- *"Software engineering principles first":* This approach and the abstract-first approach tend to share the problems of the top-down approach: without concrete examples and practical experience, you simply cannot appreciate the value of abstraction and proper software development practices.
- *"Object-oriented from day one":* Object-oriented programming is one of the best ways of organizing code and programming efforts, but it is not the only effective way. In particular, we feel that a grounding in the basics of types and algorithmic code is a prerequisite for appreciation of the design of classes and class hierarchies. We do use user-defined types (what some people would call "objects") from day one, but we don't show how to design a class until Chapter 6 and don't show a class hierarchy until Chapter 12.
- *"Just believe in magic":* This approach relies on demonstrations of powerful tools and techniques without introducing the novice to the underlying techniques and facilities. This leaves the student guessing — and usually guessing wrong — about why things are the way they are, what it costs to use them, and where they can be reasonably applied. This can lead to overrigid following of familiar patterns of work and become a barrier to further learning.

Naturally, we do not claim that these other approaches are never useful. In fact, we use several of these for specific subtopics where their strengths can be appreciated. However, as general approaches to learning programming aimed at real-world use, we reject them and apply our alternative: concrete-first and depth-first with an emphasis on concepts and techniques.

0.2.2 Programming and programming language

We teach programming first and treat our chosen programming language as secondary, as a tool. Our general approach can be used with any general-purpose programming language. Our primary aim is to help you learn general concepts, principles, and techniques. However, those cannot be appreciated in isolation. For example, details of syntax, the kinds of ideas that can be directly expressed, and tool support differ from programming language to programming language. However, many of the fundamental techniques for producing bug-free code, such as writing logically simple code (Chapters 5 and 6), establishing invariants (§9.4.3), and separating interfaces from implementation details (§9.7 and §14.1–2), vary little from programming language to programming language.

Programming and design techniques must be learned using a programming language. Design, code organization, and debugging are not skills you can acquire in the abstract. You need to write code in some programming language and gain practical experience with that. This implies that you must learn the basics of a programming language. We say "the basics" because the days when you could learn all of a major industrial language in a few weeks are gone for good. The parts of C++ we present were chosen as the subset that most directly supports the production of good code. Also, we present C++ features that you can't avoid encountering either because they are necessary for logical completeness or are common in the C++ community.

0.2.3 Portability

It is common to write C++ to run on a variety of machines. Major C++ applications run on machines we haven't ever heard of! We consider portability and the use of a variety of machine architectures and operating systems most important. Essentially every example in this book is not only ISO Standard C++, but also portable. Unless specifically stated, the code we present should work on every C++ implementation and has been tested on several machines and operating systems.

The details of how to compile, link, and run a C++ program differ from system to system. It would be tedious to mention the

details of every system and every compiler each time we need to refer to an implementation issue. In Appendix C, we give the most basic information about getting started using Visual Studio and Microsoft C++ on a Windows machine.

If you have trouble with one of the popular, but rather elaborate, IDEs (integrated development environments), we suggest you try working from the command line; it's surprisingly simple. For example, here is the full set of commands needed to compile, link, and execute a simple program consisting of two source files, **my_file1.cpp** and **my_file2.cpp**, using the GNU C++ compiler on a Unix or Linux system:

Click here to view code image

```
c++ –o my_program my_file1.cpp my_file2.cpp
./my_program
```

Yes, that really is all it takes.

0.3 Programming and computer science

Is programming all that there is to computer science? Of course not! The only reason we raise this question is that people have been known to be confused about this. We touch upon major topics from computer science, such as algorithms and data structures, but our aim is to teach programming: the design and implementation of programs. That is both more and less than most accepted notions of computer science:

- *More*, because programming involves many technical skills that are not usually considered part of any science
- *Less*, because we do not systematically present the foundation for the parts of computer science we use

The aim of this book is to be part of a course in computer science (if becoming a computer scientist is your aim), to be the foundation for the first of many courses in software construction and maintenance (if your aim is to become a programmer or a software engineer), and in general to be part of a greater whole.

We rely on computer science throughout and we emphasize principles, but we teach programming as a practical skill based on theory and experience, rather than as a science.

0.4 Creativity and problem solving

The primary aim of this book is to help you to express your ideas in code, not to teach you how to get those ideas. Along the way, we give many examples of how we can address a problem, usually through analysis of a problem followed by gradual refinement of a solution. We consider programming itself a form of problem solving: only through complete understanding of a problem and its solution can you express a correct program for it, and only through constructing and testing a program can you be certain that your understanding is complete. Thus, programming is inherently part of an effort to gain understanding. However, we aim to demonstrate this through examples, rather than through "preaching" or presentation of detailed prescriptions for problem solving.

0.5 Request for feedback

We don't think that the perfect textbook can exist; the needs of individuals differ too much for that. However, we'd like to make this book and its supporting materials as good as we can make them. For that, we need feedback; a good textbook cannot be written in isolation from its readers. Please send us reports on errors, typos, unclear text, missing explanations, etc. We'd also appreciate suggestions for better exercises, better examples, and topics to add, topics to delete, etc. Constructive comments will help future readers and we'll post errata on our support website: www.stroustrup.com/Programming.

0.6 References

Along with listing the publications mentioned in this chapter, this section also includes publications you might find helpful.

Becker, Pete, ed. *The C++ Standard.* ISO/IEC 14882:2011.

Blanchette, Jasmin, and Mark Summerfield. *C++ GUI Programming with Qt 4, Second Edition*. Prentice Hall, 2008. ISBN 0132354160.

Koenig, Andrew, and Barbara E. Moo. *Accelerated C++: Practical Programming by Example.* Addison-Wesley, 2000. ISBN 020170353X.

Meyers, Scott. *Effective C++: 55 Specific Ways to Improve Your Programs and Designs, Third Edition.* Addison-Wesley, 2005. ISBN 0321334876.

Schmidt, Douglas C., and Stephen D. Huston. *C++ Network Programming, Volume 1: Mastering Complexity with ACE and Patterns.* Addison-Wesley, 2001. ISBN 0201604647.

Schmidt, Douglas C., and Stephen D. Huston. *C++ Network Programming, Volume 2: Systematic Reuse with ACE and*

Frameworks. Addison-Wesley, 2002. ISBN 0201795256.

Stroustrup, Bjarne. *The Design and Evolution of C++*. Addison-Wesley, 1994. ISBN 0201543303.

Stroustrup, Bjarne. "Learning Standard C++ as a New Language." *C/C++ Users Journal,* May 1999.

Stroustrup, Bjarne. *The C++ Programming Language, Fourth Edition.* Addison-Wesley, 2013. ISBN 0321563840.

Stroustrup, Bjarne. *A Tour of C++*. Addison-Wesley, 2013. ISBN 0321958314.

Sutter, Herb. *Exceptional C++: 47 Engineering Puzzles, Programming Problems, and Solutions*. Addison-Wesley, 1999. ISBN 0201615622.

A more comprehensive list of references can be found in the Bibliography section at the back of the book.

0.7 Biographies

You might reasonably ask, "Who are these guys who want to teach me how to program?" So here is some biographical information. I, Bjarne Stroustrup, wrote this book, and together with Lawrence "Pete" Petersen, I designed and taught the university-level beginner's (first-year) course that was developed concurrently with the book, using drafts of the book.

Bjarne Stroustrup

I'm the designer and original implementer of the C++ programming language. I have used the language, and many other programming languages, for a wide variety of programming tasks over the last 40 years or so. I just love elegant and efficient code used in challenging applications, such as robot control, graphics, games, text analysis, and networking. I have taught design, programming, and C++ to people of essentially all abilities and interests. I'm a founding member of the ISO standards committee for C++ where I serve as the chair of the working group for language evolution.

This is my first introductory book. My other books, such as *The C++ Programming Language* and *The Design and Evolution of C++*, were written for experienced programmers.

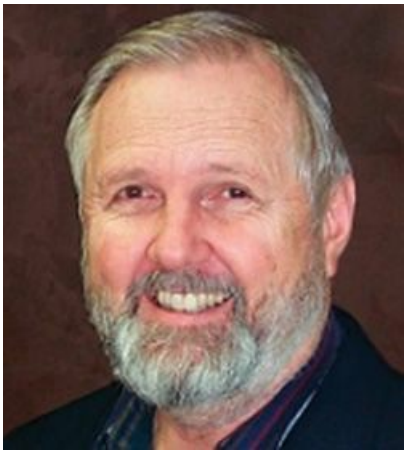
I was born into a blue-collar (working-class) family in Århus, Denmark, and got my master's degree in mathematics with computer science in my hometown university. My Ph.D. in computer science is from Cambridge University, England. I worked for AT&T for about 25 years, first in the famous Computer Science Research Center of Bell Labs — where Unix, C, C++, and so much more was invented — and later in AT&T Labs–Research.

I'm a member of the U.S. National Academy of Engineering, a Fellow of the ACM, and an IEEE Fellow. As the first computer scientist ever, I received the 2005 William Procter Prize for Scientific Achievement from Sigma Xi (the scientific research society). In 2010, I received the University of Åarhus's oldest and most prestigious honor for contributions to science by a person associated with the university, the *Rigmor og Carl Holst-Knudsens Videnskapspris*. In 2013, I was made Honorary Doctor of Computer Science from the National Research University, ITMO, St. Petersburg, Russia.

I do have a life outside work. I'm married and have two children, one a medical doctor and one a Post-doctoral Research Fellow. I read a lot (including history, science fiction, crime, and current affairs) and like most kinds of music (including classical, rock, blues, and country). Good food with friends is an essential part of life, and I enjoy visiting interesting places and people, all over the world. To be able to enjoy the good food, I run.

For more information, see my home pages: www.stroustrup.com. In particular, there you can find out how to pronounce my name.

Lawrence "Pete" Petersen



In late 2006, Pete introduced himself as follows: "I am a teacher. For almost 20 years, I have taught programming languages at Texas A&M. I have been selected by students for Teaching Excellence Awards five times and in 1996 received the Distinguished Teaching Award from the Alumni Association for the College of Engineering. I am a Fellow of the Wakonse Program for Teaching Excellence and a Fellow of the Academy for Educator Development.

"As the son of an army officer, I was raised on the move. After completing a degree in philosophy at the University of Washington, I served in the army for 22 years as a Field Artillery Officer and as a Research Analyst for Operational Testing. I taught at the Field Artillery Officers' Advanced Course at Fort Sill, Oklahoma, from 1971 to 1973. In 1979 I helped organize a Test Officers' Training Course and taught it as lead instructor at nine different locations across the United States from 1978 to 1981 and from 1985 to 1989.

"In 1991 I formed a small software company that produced management software for university departments until 1999. My interests are in teaching, designing, and programming software that real people can use. I completed master's degrees in industrial engineering at Georgia Tech and in education curriculum and instruction at Texas A&M. I also completed a master's program in microcomputers from NTS. My Ph.D. is in information and operations management from Texas A&M.

"My wife, Barbara, and I live in Bryan, Texas. We like to travel, garden, and entertain; and we spend as much time as we can with our sons and their families, and especially with our grandchildren, Angelina, Carlos, Tess, Avery, Nicholas, and Jordan."

Sadly, Pete died of lung cancer in 2007. Without him, the course would never have succeeded.

Postscript

Most chapters provide a short "postscript" that attempts to give some perspective on the information presented in the chapter. We do that with the realization that the information can be — and often is — daunting and will only be fully comprehended after doing exercises, reading further chapters (which apply the ideas of the chapter), and a later review. *Don't panic!* Relax; this is natural and expected. You won't become an expert in a day, but you can become a reasonably competent programmer as you work your way through the book. On the way, you'll encounter much information, many examples, and many techniques that lots of programmers have found stimulating and fun.

1. Computers, People, and Programming

"Specialization is for insects."

—R. A. Heinlein

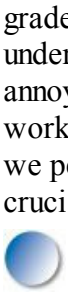
In this chapter, we present some of the things that we think make programming important, interesting, and fun. We also present a few fundamental ideas and ideals. We hope to debunk a couple of popular myths about programming and programmers. This is a chapter to skim for now and to return to later when you are struggling with some programming problem and wondering if it's all worth it.

1.1 Introduction

Like most learning, learning how to program is a chicken and egg problem: We want to get started, but we also want to know why what we are about to learn matters. We want to learn a practical skill, but also make sure it is not just a passing fad. We want to know that we are not going to waste our time, but don't want to be bored by still more hype and moralizing. For now, just read as much of this chapter as seems interesting and come back later when you feel the need to refresh your memory of why the technical details matter outside the classroom.

This chapter is a personal statement of what we find interesting and important about programming. It explains what motivates us to keep going in this field after decades. This is a chapter to read to get an idea of possible ultimate goals and an idea of what kind of person a programmer might be. A beginner's technical book inevitably contains much pretty basic stuff. In this chapter, we lift our eyes from the technical details and consider the big picture: Why is programming a worthwhile activity? What is the role of programming in our civilization? Where can a programmer make contributions to be proud of? Where does programming fit into the greater world of software development, deployment, and maintenance? When people talk about "computer science," "software engineering," "information technology," etc., where does programming fit into the picture? What does a programmer do? What skills does a good programmer have?

To a student, the most urgent reason for understanding an idea, a technique, or a chapter may be to pass a test with a good grade — but there has to be more to learning than that! To someone working in the software industry, the most urgent reason for understanding an idea, a technique, or a chapter may be to find something that can help with the current project and that will not annoy the boss who controls the next paycheck, promotions, and firings — but there has to be more to learning than that! We work best when we feel that our work in some small way makes the world a better place for people to live in. For tasks that we perform over a period of years (the "things" that professions and careers are made of), ideals and more abstract ideas are crucial.



Our civilization runs on software. Improving software and finding new uses for software are two of the ways an individual can help improve the lives of many. Programming plays an essential role in that.

1.2 Software

Good software is invisible. You can't see it, feel it, weigh it, or knock on it. *Software* is a collection of programs running on some computer. Sometimes, we can see the computer. Often, we can see only something that contains the computer, such as a telephone, a camera, a bread maker, a car, or a wind turbine. We can see what that software does. We can be annoyed or hurt if it doesn't do what it is supposed to do. We can be annoyed or hurt if what it is supposed to do doesn't suit our needs.

How many computers are there in the world? We don't know; billions at least. There may be more computers in the world than people. We need to count servers, desktop computers, laptops, tablets, smartphones, and computers embedded in "gadgets."

How many computers do you (more or less directly) use every day? There are more than 30 computers in my car, two in my cell phone, one in my MP3 player, and one in my camera. Then there is my laptop (on which the page you are reading is being written) and my desktop machine. The air-conditioning controller that keeps the summer heat and humidity at bay is a simple computer. There is one controlling the computer science department's elevator. If you use a modern television, there will be at least one computer in there somewhere. A bit of web surfing gets you into direct contact with dozens — possibly hundreds — of servers through a telecommunications system consisting of many thousands of computers — telephone switches, routers, and so on.

No, I do not drive around with 30 laptops on the backseat of my car! The point is that most computers do not look like the popular image of a computer (with a screen, a keyboard, a mouse, etc.); they are small "parts" embedded in the equipment we use. So, that car has nothing that looks like a computer, not even a screen to display maps and driving directions (though such gadgets are popular in other cars). However, its engine contains quite a few computers, doing things like fuel injection control and temperature monitoring. The power-assisted steering involves at least one computer, the radio and the security system contain some, and we suspect that even the open/close controls of the windows are computer controlled. Newer models even have computers that continuously monitor tire pressure.

How many computers do you depend on for what you do during a day? You eat; if you live in a modern city, getting the food to you is a major effort requiring minor miracles of planning, transport, and storage. The management of the distribution networks is of course computerized, as are the communication systems that stitch them all together. Modern farming is highly computerized; next to the cow barn you find computers used to monitor the herd (ages, health, milk production, etc.), farm equipment is increasingly computerized, and the number of forms required by the various branches of government can make any honest farmer cry. If something goes wrong, you can read all about it in your newspaper; of course, the articles in that paper were written on computers, set on the page by computers, and (if you still read the "dead tree edition") printed by computerized equipment — often after having been electronically transmitted to the printing plant. Books are produced in the same way. If you have to commute, the traffic flows are monitored by computers in a (usually vain) attempt to avoid traffic jams. You prefer to take the train? That train will also be computerized; some even operate without a driver, and the train's subsystems, such as announcements, braking, and ticketing, involve lots of computers. Today's entertainment industry (music, movies, television, stage shows) is among the largest users of computers. Even non-cartoon movies use (computer) animation heavily; music and photography are also digital (i.e., using computers) for both recording and delivery. Should you become ill, the tests your doctor orders will involve computers, the medical records are often computerized, and most of the medical equipment you'll encounter if you are sent to a hospital to be cured contains computers. Unless you happen to be staying in a cottage in the woods without access to any electrically powered gadgets (including light bulbs), you use energy. Oil is found, extracted, processed, and distributed through a system using computers every step along the way, from the drill bit deep in the ground to your local gas (petrol) pump. If you pay for that gas with a credit card, you again exercise a whole host of computers. It is the same story for coal, gas, solar, and wind power.

The examples so far are all "operational"; they are directly involved in what you are doing. Once removed from that is the important and interesting area of design. The clothes you wear, the telephone you talk into, and the coffee machine that dispenses your favorite brew were designed and manufactured using computers. The superior quality of modern photographic lenses and the exquisite shapes in the design of modern everyday gadgets and utensils owe almost everything to computer-based design and production methods. The craftsmen/designers/artists/engineers who design our environment have been freed from many physical constraints previously considered fundamental. If you get ill, the medicines given to cure you will have been designed using computers.

Finally, research — science itself — relies heavily on computers. The telescopes that probe the secrets of distant stars could not be designed, built, or operated without computers, and the masses of data they produce couldn't be analyzed and understood without computers. An individual biology field researcher may not be heavily computerized (unless, of course, a camera, a digital tape recorder, a telephone, etc. are used), but back in the lab, the data has to be stored, analyzed, checked against computer models, and communicated to fellow scientists. Modern chemistry and biology — including medical research — use computers to an extent undreamed of a few years ago and still unimagined by most people. The human genome was sequenced by computers. Or — let's be precise — the human genome was sequenced by humans using computers. In all of these examples, we see computers as something that enables us to do something we would have had a harder time doing

without computers.

Every one of those computers runs software. Without software, they would just be expensive lumps of silicon, metal, and plastic: doorstops, boat anchors, and space heaters. Every line of that software was written by some individual. Every one of those lines that was actually executed was minimally reasonable, if not correct. It's amazing that it all works! We are talking about billions of lines of code (program text) in hundreds of programming languages. Getting all that to work took a staggering amount of effort and involved an unimaginable number of skills. We want further improvements to essentially every service and gadget we depend on. Just think of any one service and gadget you rely on; what would you like to see improved? If nothing else, we want our services and gadgets smaller (or bigger), faster, more reliable, with more features, easier to use, with higher capacity, better looking, and cheaper. The likelihood is that the improvement you thought of requires some programming.

1.3 People

Computers are built by people for the use of people. A computer is a very generic tool; it can be used for an unimaginable range of tasks. It takes a program to make it useful to someone. In other words, a computer is just a piece of hardware until someone — some programmer — writes code for it to do something useful. We often forget about the software. Even more often, we forget about the programmer.

Hollywood and similar "popular culture" sources of disinformation have assigned largely negative images to programmers. For example, we have all seen the solitary, fat, ugly nerd with no social skills who is obsessed with video games and breaking into other people's computers. He (almost always a male) is as likely to want to destroy the world as he is to want to save it. Obviously, milder versions of such caricatures exist in real life, but in our experience they are no more frequent among software developers than they are among lawyers, police officers, car salesmen, journalists, artists, or politicians.

Think about the applications of computers you know from your own life. Were they done by a loner in a dark room? Of course not; the creation of a successful piece of software, computerized gadget, or system involves dozens, hundreds, or thousands of people performing a bewildering set of roles: for example, programmers, (program) designers, testers, animators, focus group managers, experimental psychologists, user interface designers, analysts, system administrators, customer relations people, sound engineers, project managers, quality engineers, statisticians, hardware interface engineers, requirements engineers, safety officers, mathematicians, sales support personnel, troubleshooters, network designers, methodologists, software tools managers, software librarians, etc. The range of roles is huge and made even more bewildering by the titles varying from organization to organization: one organization's "engineer" may be another organization's "programmer" and yet another organization's "developer," "member of technical staff," or "architect." There are even organizations that let their employees pick their own titles. Not all of these roles directly involve programming. However, we have personally seen examples of people performing each of the roles mentioned while reading or writing code as an essential part of their job. Additionally, a programmer (performing any of these roles, and more) may over a short period of time interact with a wide range of people from application areas, such as biologists, engine designers, lawyers, car salesmen, medical researchers, historians, geologists, astronauts, airplane engineers, lumberyard managers, rocket scientists, bowling alley builders, journalists, and animators (yes, this is a list drawn from personal experience). Someone may also be a programmer at times and fill non-programming roles at other stages of a professional career.

The myth of a programmer being isolated is just that: a myth. People who like to work on their own choose areas of work where that is most feasible and usually complain bitterly about the number of "interruptions" and meetings. People who prefer to interact with other people have an easier time because modern software development is a team activity. The implication is that social and communication skills are essential and valued far more than the stereotypes indicate. On a short list of highly desirable skills for a programmer (however you realistically define *programmer*), you find the ability to communicate well — with people from a wide variety of backgrounds — informally, in meetings, in writing, and in formal presentations. We are convinced that until you have completed a team project or two, you have no idea of what programming is and whether you really like it. Among the things we like about programming are all the nice and interesting people we meet and the variety of places we get to visit as part of our professional lives.

One implication of all this is that people with a wide variety of skills, interests, and work habits are essential for producing good software. Our quality of life depends on those people — sometimes even our life itself. No one person could fill all the roles we mention here; no sensible person would want every role. The point is that you have a wider choice than you could possibly imagine; not that you have to make any particular choice. As an individual you will "drift" toward areas of work that match your skills, talents, and interests.

We talk about "programmers" and "programming," but obviously programming is only part of the overall picture. The people who design a ship or a cell phone don't think of themselves as programmers. Programming is an important part of

software development, but not all there is to software development. Similarly, for most products, software development is an important part of product development, but not all there is to product development.

We do not assume that you — our reader — want to become a professional programmer and spend the rest of your working life writing code. Even the best programmers — especially the *best* programmers — spend most of their time *not* writing code. Understanding problems takes serious time and often requires significant intellectual effort. That intellectual challenge is what many programmers refer to when they say that programming is interesting. Many of the best programmers also have degrees in subjects not usually considered part of computer science. For example, if you work on software for genomic research, you will be much more effective if you understand some molecular biology. If you work on programs for analyzing medieval literature, you could be much better off reading a bit of that literature and maybe even knowing one or more of the relevant languages. In particular, a person with an "all I care about is computers and programming" attitude will be incapable of interacting with his or her non-programmer colleagues. Such a person will not only miss out on the best parts of human interactions (i.e., life) but also be a bad software developer.

So, what do we assume? Programming is an intellectually challenging set of skills that are part of many important and interesting technical disciplines. In addition, programming is an essential part of our world, so not knowing the basics of programming is like not knowing the basics of physics, history, biology, or literature. Someone totally ignorant of programming is reduced to believing in magic and is dangerous in many technical roles. If you read Dilbert, think of the pointy-haired boss as the kind of manager you don't want to meet or (far worse) become. In addition, programming can be fun.

But what do we assume you might use programming for? Maybe you will use programming as a key tool in your further studies and work without becoming a professional programmer. Maybe you will interact with other people professionally and personally in ways where a basic knowledge of programming will be an advantage, maybe as a designer, writer, manager, or scientist. Maybe you will do programming at a professional level as part of your studies or work. Even if you do become a professional programmer it is unlikely that you will do nothing but programming.

You might become an engineer focusing on computers or a computer scientist, but even then you will not "program all the time." Programming is a way of presenting ideas in code — a way of aiding problem solving. It is nothing — absolutely a waste of time — unless you have ideas that are worth presenting and problems worth solving.

This is a book about programming and we have promised to help you learn how to program, so why do we emphasize non-programming subjects and the limited role of programming? A good programmer understands the role of code and programming technique in a project. A good programmer is (at most times) a good team player and tries hard to understand how the code and its production best support the overall project. For example, imagine that I worked on a new MP3 player (maybe to be part of a smartphone or a tablet) and all that I cared about was the beauty of my code and the number of neat features I could provide. I would probably insist on the largest, most powerful computer to run my code. I might disdain the theory of sound encoding because it is "not programming." I would stay in my lab, rather than go out to meet potential users, who undoubtedly would have bad tastes in music anyway and would not appreciate the latest advances in GUI (graphical user interface) programming. The likely result would be disaster for the project. A bigger computer would mean a costlier MP3 player and most likely a shorter battery life. Encoding is an essential part of handling music digitally, so failing to pay attention to advances in encoding techniques could lead to increased memory requirements for each song (encodings differ by as much as 100% for the same-quality output). A disregard for users' preferences — however odd and archaic they may seem to you — typically leads to the users choosing some other product. An essential part of writing a good program is to understand the needs of the users and the constraints that those needs place on the implementation (i.e., the code). To complete this caricature of a bad programmer, we just have to add a tendency to deliver late because of an obsession with details and an excessive confidence in the correctness of lightly tested code. We encourage you to become a good programmer, with a broad view of what it takes to produce good software. That's where both the value to society and the keys to personal satisfaction lie.

1.4 Computer science

Even by the broadest definition, programming is best seen as a part of something greater. We can see it as a subdiscipline of computer science, computer engineering, software engineering, information technology, or any other software-related discipline. We see programming as an enabling technology for those computer and information fields of science and engineering, as well as for physics, biology, medicine, history, literature, and any other academic or research field.

Consider computer science. A 1995 U.S. government "blue book" defines it like this: "The systematic study of computing systems and computation. The body of knowledge resulting from this discipline contains theories for understanding computing systems and methods; design methodology, algorithms, and tools; methods for the testing of concepts; methods of analysis and verification; and knowledge representation and implementation." As we would expect, the Wikipedia entry is less formal: "Computer science, or computing science, is the study of the theoretical foundations of information and computation and their

implementation and application in computer systems. Computer science has many sub-fields; some emphasize the computation of specific results (such as computer graphics), while others (such as computational complexity theory) relate to properties of computational problems. Still others focus on the challenges in implementing computations. For example, programming language theory studies approaches to describing computations, while computer programming applies specific programming languages to solve specific computational problems."

Programming is a tool; it is a fundamental tool for expressing solutions to fundamental and practical problems so that they can be tested, improved through experiment, and used. Programming is where ideas and theories meet reality. This is where computer science can become an experimental discipline, rather than pure theory, and impact the world. In this context, as in many others, it is essential that programming is an expression of well-tried practices as well as the theories. It must not degenerate into mere hacking: just get some code written, any old way that meets an immediate need.

1.5 Computers are everywhere

Nobody knows everything there is to know about computers or software. This section just gives you a few examples. Maybe you'll see something you like. At least you might be convinced that the scope of computer use — and through that, programming — is far larger than any individual can fully grasp.

Most people think of a computer as a small gray box attached to a screen and a keyboard. Such computers tend to be good at games, messaging and email, and playing music. Other computers, called laptops, are used on planes by bored businessmen to look at spreadsheets, play games, and watch videos. This caricature is just the tip of the iceberg. Most computers work out of our sight and are part of the systems that keep our civilization going. Some fill rooms; others are smaller than a small coin. Many of the most interesting computers don't directly interact with a human through a keyboard, mouse, or similar gadget.

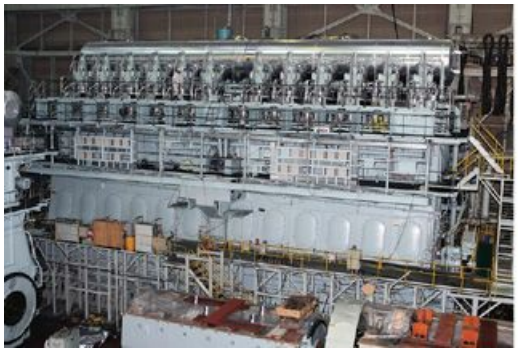
1.5.1 Screens and no screens

The idea of a computer as a fairly large rectangular box with a screen and a keyboard is common and often hard to shake off. However, consider these two computers:

Both of these "gadgets" (which happen to be watches) are primarily computers. In fact, we conjecture that they are essentially the same model computer with different I/O (input/output) systems. The left one drives a small screen (similar to the screens on conventional computers, but smaller) and the second drives little electric motors controlling traditional clock hands and a disk of numbers for day-of-month readout. Their input systems are the four buttons (more easily seen on the right-hand watch) and a radio receiver, used for synchronization with very high-precision "atomic" clocks. Most of the programs controlling these two computers are shared between them.

1.5.2 Shipping

These two photos show a large marine diesel engine and the kind of huge ship that it may power:



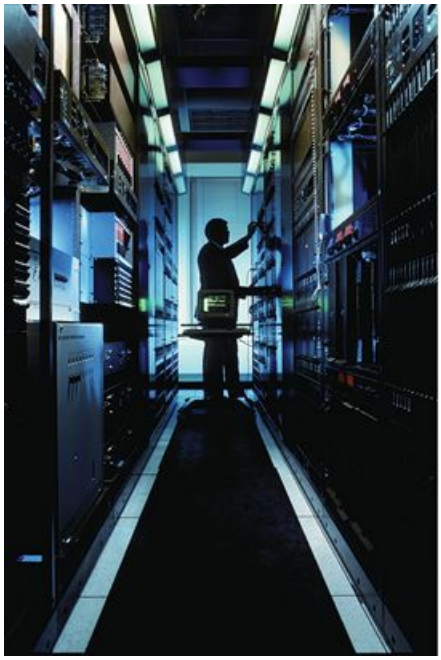

Consider where computers and software play key roles here:

- *Design:* Of course, the ship and the engine were both designed using computers. The list of uses is almost endless and includes architectural and engineering drawings, general calculations, visualization of spaces and parts, and simulations of the performance of parts.
- *Construction:* A modern shipyard is heavily computerized. The assembly of a ship is carefully planned using computers, and the work is guided by computers. Welding is done by robots. In particular, a modern double-hulled tanker couldn't be built without little welding robots to do the welding from within the space between the hulls. There just isn't room for a human in there. Cutting steel plates for a ship was one of the world's first CAD/CAM (computer-aided design and computer-aided manufacture) applications.
- *The engine:* The engine has electronic fuel injection and is controlled by a few dozen computers. For a 100,000-horsepower engine (like the one in the photo), that's a nontrivial task. For example, the engine management computers continuously adjust fuel mix to minimize the pollution that would result from a badly tuned engine. Many of the pumps associated with the engine (and other parts of the ship) are themselves computerized.
- *Management:* Ships sail where there is cargo to pick up and to deliver. The scheduling of fleets of ships is a continuing process (computerized, of course) so that routings change with the weather, with supply and demand, and with space and loading capacity of harbors. There are even websites where you can watch the position of major merchant vessels at any time. The ship in the photo happens to be a container vessel (one of the largest such in the world; 397m long and 56m wide), but other kinds of large modern ships are managed in similar ways.
- *Monitoring:* An oceangoing ship is largely autonomous; that is, its crew can handle most contingencies likely to arise before the next port. However, they are also part of a globe-spanning network. The crew has access to reasonably accurate weather information (from and through — computerized — satellites). They have a GPS (global positioning system) and computer-controlled and computer-enhanced radar. If the crew needs a rest, most systems (including the engine, radar, etc.) can be monitored (via satellite) from a shipping-line control room. If anything unusual is spotted, or if the connection "back home" is broken, the crew is notified.

Consider the implication of a failure of one of the hundreds of computers explicitly mentioned or implied in this brief description. Chapter 25 ("Embedded Systems Programming") examines this in slightly more detail. Writing code for a modern ship is a skilled and interesting activity. It is also useful. The cost of sea transport is really amazingly low. You appreciate that when you buy something that wasn't manufactured locally. Sea transport has always been cheaper than land transport; these days one of the reasons is serious use of computers and information.

1.5.3 Telecommunications

These two photos show a telephone switch and a telephone (that also happens to be a camera, an MP3 player, an FM radio, a web browser, and much more):

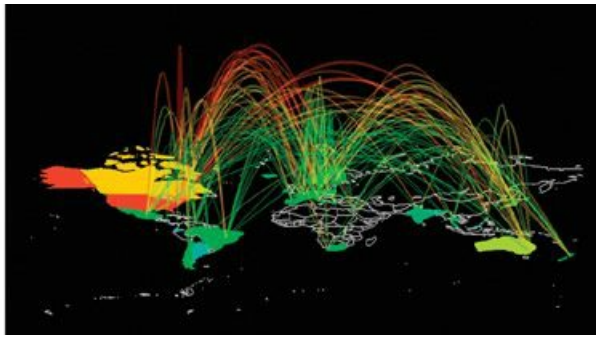


Consider where computers and software play key roles here. You pick up a telephone and "dial," the person you dialed answers, and you talk. Or maybe you get to leave a voicemail, or maybe you send a photo from your phone camera, or maybe you send a text message (hit Send and let the phone do the dialing). Obviously the phone is a computer. This is especially obvious if the phone (like most mobile phones) has a screen and allows more than traditional "plain old telephone services," such as web browsing. Actually, such phones tend to contain several computers: one to manage the screen, one to talk to the phone system, and maybe more.

The part of the phone that manages the screen, does web browsing, etc. is probably the most familiar to computer users: it just runs a graphical user interface to "all the usual stuff." What is unknown to and largely unsuspected by most users is the huge system that the little phone talks to while doing its job. I dial a number in Texas, but you are on vacation in New York City, yet within seconds your phone rings and I hear your "Hello!" over the roar of city traffic. Many phones can perform that trick for essentially any two locations on earth and we just take it for granted. How did my phone find yours? How is the sound transmitted? How is the sound encoded into data packets? The answer could fill many books much thicker than this one, but it involves a combination of hardware and software on hundreds of computers scattered over the geographical area in question. If you are unlucky, a few telecommunications satellites (themselves computerized systems) are also involved — "unlucky" because we cannot perfectly compensate for the 20,000-mile detour out into space; the speed of light (and therefore the speed of your voice) is finite (light fiber cables are much better: shorter, faster, and carrying much more data). Most of this works remarkably well; the backbone telecommunications systems are 99.9999% reliable (for example, 20 minutes of downtime in 20 years — that's 20/20*365*24*60). The trouble we have tends to be in the communications between our mobile phone and the nearest main telephone switch.

There is software for connecting the phones, for chopping our spoken words into data packets to be sent over wires and radio links, for routing those messages, for recovering from all kinds of failures, for continuously monitoring the quality and reliability of the services, and of course for billing. Even keeping track of all the physical pieces of the system requires serious amounts of clever software: What talks to what? What parts go into a new system? When do you need to do some preventive maintenance?

Arguably the backbone telecommunications system of the world, consisting of semi-independent but interconnected systems, is the largest and most complicated man-made artifact. To make things a bit more real: remember, this is not just boring old telephony with a few new bells and whistles. The various infrastructures have merged. They are also what the internet (the web) runs on, what our banking and trading systems run on, and what carry our television programs to the broadcasting stations. So, we can add another couple of photos to illustrate telecommunications:

The room is the "trading floor" of the American stock exchange on New York's Wall Street and the map is a representation of parts of the internet backbones (a complete map would be too messy to be useful).

As it happens, we also like digital photography and the use of computers to draw specialized maps to visualize knowledge.

1.5.4 Medicine

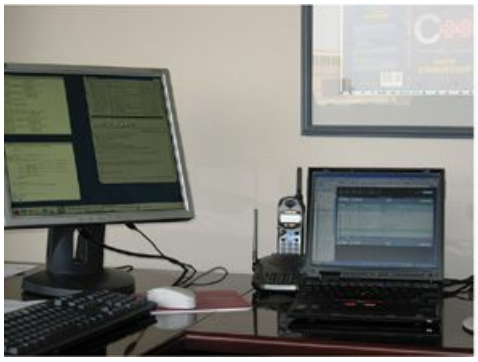
These two photos show a CAT (computed axial tomography) scanner and an operating theater for computer-aided surgery (also called "robot-assisted surgery" or "robotic surgery"):


Consider where computers and software play key roles here. The scanners basically are computers; the pulses they send out are controlled by a computer, and the readings are nothing but gibberish until quite sophisticated algorithms are applied to convert them to something we recognize as a (three-dimensional) image of the relevant part of a human body. To do computerized surgery, we must go several steps further. A wide variety of imaging techniques are used to let the surgeon see the inside of the patient, to see the point of surgery with significant enlargement or in better light than would otherwise be possible. With the aid of a computer a surgeon can use tools that are too fine for a human hand to hold or in a place where a human hand could not reach without unnecessary cutting. The use of minimally invasive surgery (laparoscopic surgery) is a simple example of this that has minimized the pain and recovery time for millions of people. The computer can also help steady the surgeon's "hand" to allow for more delicate work than would otherwise be possible. Finally, a "robotic" system can be operated remotely, thus making it possible for a doctor to help someone remotely (over the internet). The computers and programming involved are mind-boggling, complex, and interesting. The user interface, equipment control, and imaging challenges alone will keep thousands of researchers, engineers, and programmers busy for decades.

We heard of a discussion among a large group of medical doctors about which new tool had provided the most help to them in their work: The CAT scanner? The MRI scanner? The automated blood analysis machines? The high-resolution ultrasound machines? PDAs? After some discussion, a surprising "winner" of this "competition" emerged: instant access to patient records. Knowing the medical history of a patient (earlier illnesses, medicines tried earlier, allergies, hereditary problems, general health, current medication, etc.) simplifies the problem of diagnosis and minimizes the chance of mistakes.

1.5.5 Information

These two photos show an ordinary PC (well, two) and part of a server farm:




We have focused on "gadgets" for the usual reason: you cannot see, feel, or hear software. We cannot present you with a photograph of a neat program, so we show you a "gadget" that runs one. However, much software deals directly with "information." So let's consider "ordinary uses" of "ordinary computers" running "ordinary software."

A "server farm" is a collection of computers providing web services. Organizations running state-of-the-art server farms (such as Google, Amazon, and Microsoft) are somewhat close-mouthed about the details of their servers, and the specifications of server farms change constantly (so most of the information you find on the web is outdated). However, the specifications are amazing and should convince anyone that there is more to programming than simply computing a few numbers on a laptop:

- Google uses about a million servers (each more powerful than your laptop) in 25 to 50 "data centers."
- Such a data center is housed in a warehouse that might measure 60m*100m (that's about 200ft*330ft) or more.
- In 2011, the *New York Times* reported that Google's data centers draw about 260 million watts continuously (about the same amount of energy as Las Vegas).
- Assume a server machine to be a 3GHz quad-core with 24GB of main memory. That would imply about $12*10^{15}$Hz of compute power (about 12,000,000,000,000,000 instructions per second) with $24*10^{15}$ bytes of main memory (about 24,000,000,000,000,000 8-bit bytes), and maybe 4TB of disk per server, giving $4*10^{18}$ bytes of storage.

We may be underestimating the amounts, and by the time you read this, we almost certainly are. In particular, efforts to minimize energy usage seem to be driving machine architectures toward more processors per server and more cores per processor. A GB is a gigabyte, that is, about 10^9 characters. A TB, a terabyte, is about 1000GB, that is, about 10^{12} characters. A PB, a petabyte (that is, 10^{15} bytes), is becoming a more common measure. This is a pretty extreme example, but every major company runs programs on the web to interact with its users/customers. Examples are Amazon (book and other sales), Amadeus (airline ticketing and automobile rental), and eBay (online auctions). Millions of companies, organizations, and individuals also have a presence on the web. Most don't run their own software, but many do and much of that is not trivial.

The other, and more traditional, massive computing effort involves accounting, order processing, payroll, record keeping, billing, inventory management, personnel records, student records, patient records, etc. — the records that essentially every organization (commercial and noncommercial, governmental and private) keeps. These records are the backbone of their respective organizations. As a computing effort, processing such records seems simple: mostly some information (records) is just stored and retrieved and very little is done to it. Examples include

- Is my 12:30 flight to Chicago still on time?
- Has Gilbert Sullivan had the measles?
- Has the coffeemaker that Juan Valdez ordered been shipped?
- What kind of kitchen chair did Jack Sprat buy in 1996 (or so)?
- How many phone calls originated from the 212 area code in August of 2012?
- What was the number of coffeepots sold in January and for what total price?

The sheer scale of the databases involved makes these systems highly complex. To that add the need to respond quickly (often in less than two seconds for individual queries) and to be correct (at least most of the time). These days, it is not uncommon for people to talk about terabytes of data (a byte is the amount of memory needed to hold an ordinary character). That's traditional "data processing" and it is merging with "the web" because most access to the databases is now through web interfaces.

This kind of computer use is often referred to as *information processing*. It focuses on data — often lots of data. This leads to challenges in the organization and transmission of data and lots of interesting work on how to present vast amounts of data in a comprehensible form: "user interface" is a very important aspect of handling data. For example, think of analyzing a work of older literature (say, Chaucer's *Canterbury Tales* or Cervantes' *Don Quixote*) to figure out what the author actually wrote by comparing dozens of versions. We need to search through the texts with a variety of criteria supplied by the person doing the analysis and to display the results in a way that aids the discovery of salient points. Thinking of text analysis, publishing comes to mind: today, just about every article, book, brochure, newspaper, etc. is produced on a computer. Designing software to

support that well is for most people still a problem that lacks a really good solution.

1.5.6 A vertical view

It is sometimes claimed that a paleontologist can reconstruct a complete dinosaur and describe its lifestyle and natural environment from studying a single small bone. That may be an exaggeration, but there is something to the idea of looking at a simple artifact and thinking about what it implies. Consider this photo showing the landscape of Mars taken by a camera on one of NASA's Mars Rovers:

If you want to do "rocket science," becoming a good programmer is one way. The various space programs employ lots of software designers, especially ones who can also understand some of the physics, math, electrical engineering, mechanical engineering, medical engineering, etc. that underlie the manned and unmanned space programs. Getting those two Rovers to drive around on Mars for years is one of the greatest technological triumphs of our civilization. One (*Spirit*) sent data back for six years and the other (*Opportunity*) is still working at the time of writing and will have its tenth anniversary on Mars in January 2014. Their estimated design life was three months.

The photo was transmitted to earth through a communication channel with a 25-minute transmission delay each way; there is a lot of clever programming and advanced math to make sure that the picture is transmitted using the minimal number of bits without losing any of them. On earth, the photo is then rendered using algorithms to restore color and minimize distortion due to the optics and electronic sensors.

The control programs for the Mars Rovers are of course programs — the Rovers drive autonomously for 24 hours at a time and follow instructions sent from earth the day before. The transmission is managed by programs.

The operating systems used for the various computers involved in the Rovers, the transmission, and the photo reconstruction are programs, as are the applications used to write this chapter. The computers on which these programs run are designed and produced using CAD/CAM (computer-aided design and computer-aided manufacture) programs. The chips that go into those computers are produced on computerized assembly lines constructed using precision tools, and those tools also use computers (and software) in their design and manufacture. The quality control for those long construction processes involves serious computation. All that code was written by humans in a high-level programming language and translated into machine code by a compiler, which is itself such a program. Many of these programs interact with users using GUIs and exchange data using input/output streams.

Finally, a lot of programming goes into image processing (including the processing of the photos from the Mars Rovers), animation, and photo editing (there are versions of the Rover photos floating around on the web featuring "Martians").

1.5.7 So what?

What do all these "fancy and complicated" applications and software systems have to do with learning programming and using C++? The connection is simply that many programmers do get to work on projects like these. These are the kinds of things that good programming can help achieve. Also, every example used in this chapter involved C++ and at least some of the techniques we describe in this book. Yes, there are C++ programs in MP3 players, in ships, in wind turbines, on Mars, and in the human genome project. For more applications using C++, see www.stroustrup.com/applications.html.

1.6 Ideals for programmers

What do we want from our programs? What do we want in general, as opposed to a particular feature of a particular program? We want *correctness* and as part of that, *reliability*. If the program doesn't do what it is supposed to do, and do so in a way so that we can rely on it, it is at best a serious nuisance, at worst a danger. We want it to be *well designed* so that it addresses a real need well; it doesn't really matter that a program is correct if what it does is irrelevant to us or if it correctly does

something in a way that annoys us. We also want it to be *affordable*; I might prefer a Rolls-Royce or an executive jet to my usual forms of transport, but unless I'm a zillionaire, cost will enter into my choices.

These are aspects of software (gadgets, systems) that can be appreciated from the outside, by non-programmers. They must be ideals for programmers and we must keep them in mind at all times, especially in the early phases of development, if we want to produce successful software. In addition, we must concern ourselves with ideals related to the code itself: our code must be *maintainable*; that is, its structure must be such that someone who didn't write it can understand it and make changes. A successful program "lives" for a long time (often for decades) and will be changed again and again. For example, it will be moved to new hardware, it will have new features added, it will be modified to use new I/O facilities (screens, video, sound), to interact using new natural languages, etc. Only a failed program will never be modified. To be maintainable, a program must be simple relative to its requirements, and the code must directly represent the ideas expressed. Complexity — the enemy of simplicity and maintainability — can be intrinsic to a problem (in that case we just have to deal with it), but it can also arise from poor expression of ideas in code. We must try to avoid that through good coding style — style matters!

This doesn't sound too difficult, but it is. Why? Programming is fundamentally simple: just tell the machine what it is supposed to do. So why can programming be most challenging? Computers are fundamentally simple; they can just do a few operations, such as adding two numbers and choosing the next instruction to execute based on a comparison of two numbers. The problem is that we don't want computers to do simple things. We want "the machine" to do things that are difficult enough for us to want help with them, but computers are nitpicking, unforgiving, dumb beasts. Furthermore, the world is more complex than we'd like to believe, so we don't really know the implications of what we request. We just want a program to "do something like this" and don't want to be bothered with technical details. We also tend to assume "common sense." Unfortunately, common sense isn't all that common among humans and is totally absent in computers (though some really well-designed programs can imitate it in specific, well-understood cases).

This line of thinking leads to the idea that "programming is understanding": when you can program a task, you understand it. Conversely, when you understand a task thoroughly, you can write a program to do it. In other words, we can see programming as part of an effort to thoroughly understand a topic. A program is a precise representation of our understanding of a topic.

When you program, you spend significant time trying to understand the task you are trying to automate.

We can describe the process of developing a program as having four stages:

- *Analysis:* What's the problem? What does the user want? What does the user need? What can the user afford? What kind of reliability do we need?
- *Design:* How do we solve the problem? What should be the overall structure of the system? Which parts does it consist of? How do those parts communicate with each other? How does the system communicate with its users?
- *Programming:* Express the solution to the problem (the design) in code. Write the code in a way that meets all constraints (time, space, money, reliability, and so on). Make sure that the code is correct and maintainable.
- *Testing:* Make sure the system works correctly under all circumstances required by systematically trying it out.

Programming plus testing is often called *implementation*. Obviously, this simple split of software development into four parts is a simplification. Thick books have been written on each of these four topics and more books still about how they relate to each other. One important thing to note is that these stages of development are not independent and do not occur strictly in sequence. We typically start with analysis, but feedback from testing can help improve the programming; problems with getting the program working may indicate a problem with the design; and working with the design may suggest aspects of the problem that hitherto had been overlooked in the analysis. Actually using the system typically exposes weaknesses of the analysis.

The crucial concept here is *feedback*. We learn from experience and modify our behavior based on what we learn. That's essential for effective software development. For any large project, we don't know everything there is to know about the problem and its solution before we start. We can try out ideas and get feedback by programming, but in the earlier stages of development it is easier (and faster) to get feedback by writing down design ideas, trying out those design ideas, and using scenarios on friends. The best design tool we know of is a blackboard (use a whiteboard instead if you prefer chemical smells over chalk dust). Never design alone if you can avoid it! Don't start coding before you have tried out your ideas by explaining them to someone. Discuss designs and programming techniques with friends, colleagues, potential users, and so on before you

head for the keyboard. It is amazing how much you can learn from simply trying to articulate an idea. After all, a program is nothing more than an expression (in code) of some ideas.

Similarly, when you get stuck implementing a program, look up from the keyboard. Think about the problem itself, rather than your incomplete solution. Talk with someone: explain what you want to do and why it doesn't work. It's amazing how often you find the solution just by carefully explaining the problem to someone. Don't debug (find program errors) alone if you don't have to!

The focus of this book is implementation, and especially programming. We do not teach "problem solving" beyond giving you plenty of examples of problems and their solutions. Much of problem solving is recognizing a known problem and applying a known solution technique. Only when most subproblems are handled this way will you find the time to indulge in exciting and creative "out-of-the-box thinking." So, we focus on showing how to express ideas clearly in code.

Direct expression of ideas in code is a fundamental ideal of programming. That's really pretty obvious, but so far we are a bit short of good examples. We'll come back to this, repeatedly. When we want an integer in our code, we store it in an **int**, which provides the basic integer operations. When we want a string of characters, we store it in a **string**, which provides the most basic text manipulation operations. At the most fundamental level, the ideal is that when we have an idea, a concept, an entity, something we think of as a "thing," something we can draw on our whiteboard, something we can refer to in our discussions, something our (non–computer science) textbook talks about, then we want that something to exist in our program as a named entity (a type) providing the operations we think appropriate for it. If we want to do math, we want a **complex** type for complex numbers and a **Matrix** type for linear algebra. If we want to do graphics, we want a **Shape** type, a **Circle** type, a **Color** type, and a **Dialog_box**. When we want to deal with streams of data, say from a temperature sensor, we want an **istream** type (**i** for input). Obviously, every such type should provide the appropriate operations and only the appropriate operations. These are just a few examples from this book. Beyond that, we offer tools and techniques for you to build your own types to directly represent whatever concepts you want in your program.

Programming is part practical, part theoretical. If you are just practical, you will produce non-scalable, unmaintainable hacks. If you are just theoretical, you will produce unusable (or unaffordable) toys.

For a different kind of view of the ideals of programming and a few people who have contributed in major ways to software through work with programming languages, see Chapter 22, "Ideals and History."

Review

Review questions are intended to point you to the key ideas explained in a chapter. One way to look at them is as a complement to the exercises: the exercises focus on the practical aspects of programming, whereas the review questions try to help you articulate the ideas and concepts. In that, they resemble good interview questions.

1. What is software?
2. Why is software important?
3. Where is software important?
4. What could go wrong if some software fails? List some examples.
5. Where does software play an important role? List some examples.
6. What are some jobs related to software development? List some.
7. What's the difference between computer science and programming?
8. Where in the design, construction, and use of a ship is software used?
9. What is a server farm?
10. What kinds of queries do you ask online? List some.
11. What are some uses of software in science? List some.
12. What are some uses of software in medicine? List some.
13. What are some uses of software in entertainment? List some.
14. What general properties do we expect from good software?
15. What does a software developer look like?
16. What are the stages of software development?
17. Why can software development be difficult? List some reasons.

18. What are some uses of software that make your life easier?

19. What are some uses of software that make your life more difficult?

Terms

These terms present the basic vocabulary of programming and of C++. If you want to understand what people say about programming topics and to articulate your own ideas, you should know what each means.

affordability
analysis
blackboard
CAD/CAM
communication
correctness
customer
design
feedback
GUI
ideals
implementation
programmer
programming
software
stereotype
testing
user

Exercises

1. Pick an activity you do most days (such as going to class, eating dinner, or watching television). Make a list of ways computers are directly or indirectly involved.

2. Pick a profession, preferably one that you have some interest in or some knowledge of. Make a list of activities done by people in that profession that involve computers.

3. Swap your list from exercise 2 with a friend who picked a different profession and improve his or her list. When you have both done that, compare your results. Remember: There is no perfect solution to an open-ended exercise; improvements are always possible.

4. From your own experience, describe an activity that would not have been possible without computers.

5. Make a list of programs (software applications) that you have directly used. List only examples where you obviously interact with a program (such as when selecting a new song on an MP3 player) and not cases where there just might happen to be a computer involved (such as turning the steering wheel of your car).

6. Make a list of ten activities that people do that do not involve computers in any way, even indirectly. This may be harder than you think!

7. Identify five tasks for which computers are not used today, but for which you think they will be used at some time in the future. Write a few sentences to elaborate on each one that you choose.

8. Write an explanation (at least 100 words, but fewer than 500) of why you would like to be a computer programmer. If, on the other hand, you are convinced that you would not like to be a programmer, explain that. In either case, present well-thought-out, logical arguments.

9. Write an explanation (at least 100 words, but fewer than 500) of what role other than programmer you'd like to play in the computer industry (independently of whether "programmer" is your first choice).

10. Do you think computers will ever develop to be conscious, thinking beings, capable of competing with humans? Write a short paragraph (at least 100 words) supporting your position.

11. List some characteristics that most successful programmers share. Then list some characteristics that programmers are

popularly assumed to have.

12. Identify at least five kinds of applications for computer programs mentioned in this chapter and pick the one that you find the most interesting and that you would most likely want to participate in someday. Write a short paragraph (at least 100 words) explaining why you chose the one you did.

13. How much memory would it take to store (a) this page of text, (b) this chapter, (c) all of Shakespeare's work? Assume one byte of memory holds one character and just try to be precise to about 20%.

14. How much memory does your computer have? Main memory? Disk?

Postscript

Our civilization runs on software. Software is an area of unsurpassed diversity and opportunities for interesting, socially useful, and profitable work. When you approach software, do it in a principled and serious manner: you want to be part of the solution, not add to the problems.

We are obviously in awe of the range of software that permeates our technological civilization. Not all applications of software do good, of course, but that is another story. Here we wanted to emphasize how pervasive software is and how much of what we rely on in our daily lives depends on software. It was all written by people like us. All the scientists, mathematicians, engineers, programmers, etc. who built the software briefly mentioned here started like you are starting.

Now, let's get back to the down-to-earth business of learning the technical skills needed to program. If you start wondering if it is worth all your hard work (most thoughtful people wonder about that sometime), come back and reread this chapter, the Preface, and bits of Chapter 0 ("Notes to the Reader"). If you start wondering if you can handle it all, remember that millions have succeeded in becoming competent programmers, designers, software engineers, etc. You can, too.

Part I: The Basics

2. Hello, World!

"Programming is learned by writing programs."

—Brian Kernighan

Here, we present the simplest C++ program that actually does anything. The purpose of writing this program is to

- Let you try your programming environment
- Give you a first feel of how you can get a computer to do things for you

Thus, we present the notion of a program, the idea of translating a program from human-readable form to machine instructions using a compiler, and finally executing those machine instructions.

2.1 Programs

To get a computer to do something, you (or someone else) have to tell it exactly — in excruciating detail — what to do. Such a description of "what to do" is called a *program*, and *programming* is the activity of writing and testing such programs.

In a sense, we have all programmed before. After all, we have given descriptions of tasks to be done, such as "how to drive to the nearest cinema," "how to find the upstairs bathroom," and "how to heat a meal in the microwave." The difference between such descriptions and programs is one of degree of precision: humans tend to compensate for poor instructions by using common sense, but computers don't. For example, "turn right in the corridor, up the stairs, it'll be on your left" is probably a fine description of how to get to the upstairs bathroom. However, when you look at those simple instructions, you'll find the grammar sloppy and the instructions incomplete. A human easily compensates. For example, assume that you are sitting at the table and ask for directions to the bathroom. You don't need to be told to get up from your chair to get to the corridor, somehow walk around (and not across or under) the table, not to step on the cat, etc. You'll not have to be told not to bring your knife and fork or to remember to switch on the light so that you can see the stairs. Opening the door to the bathroom before entering is probably also something you don't have to be told.

In contrast, computers are *really* dumb. They have to have everything described precisely and in detail. Consider again "turn right in the corridor, up the stairs, it'll be on your left." Where is the corridor? What's a corridor? What is "turn right"? What stairs? How do I go up stairs? (One step at a time? Two steps? Slide up the banister?) What is on my left? When will it be on my left? To be able to describe "things" precisely for a computer, we need a precisely defined language with a specific grammar (English is far too loosely structured for that) and a well-defined vocabulary for the kinds of actions we want performed. Such a language is called a *programming language*, and C++ is a programming language designed for a wide selection of programming tasks.

If you want greater philosophical detail about computers, programs, and programming, (re)read Chapter 1. Here, let's have a look at some code, starting with a very simple program and the tools and techniques you need to get it to run.

2.2 The classic first program

Here is a version of the classic first program. It writes "Hello, World!" to your screen:

Click here to view code image

```
// This program outputs the message "Hello, World!" to the monitor

#include "std_lib_facilities.h"

int main()          // C++ programs start by executing the function main
{
        cout << "Hello, World!\n";      // output "Hello, World!"
        return 0;
}
```