Foreword

This is a delightful book on data structures that are both time and space efficient. Space as well as time efficiency is crucial in modern information systems. Even if we have extra space somewhere, it is unlikely to be close to the processors. The space used by most such systems is overwhelmingly for structural indexing, such as B-trees, hash tables, and various cross-references, rather than for "raw data." Indeed data, such as text, take far too much space in raw form and must be compressed. A system that keeps both data and indices in a compact form has a major advantage.

Hence the title of the book. Gonzalo Navarro uses the term "compact data structures" to describe a newly emerging research area. It has developed from two distinct but interrelated topics. The older is that of text compression, dating back to the work of Shannon, Fano, and Huffman (among others) in the late 1940s and early 1950s (although text compression as such was not their main concern). Through the last half of the 20th century, as the size of the text to be processed increased and computing platforms became more powerful, algorithmics and information theory became much more sophisticated. The goal of data compression, at least until the year 2000 or so, simply meant compressing information as well as possible and then decompressing each time it was needed. A hallmark of compact data structures is working with text in compressed form saving both decompression time and space. The newer contributing area evolved in the 1990s after the work of Jacobson and is generally referred to as "succinct data structures." The idea is to represent a combinatorial object, such as a graph, tree, or sparse bit vector, in a number of bits that differs from the information theory lower bound by only a lower order term. So, for example, a binary tree on n nodes takes only $2n + o(n)$ bits. The trick is to perform the necessary operations, e.g., find child, parent, or subtree size, in constant time.

Compact data structures take into account both "data" and "structures" and are a little more tolerant of "best effort" than one might be with exact details of information theoretic lower bounds. Here the subtitle, "A Practical Approach," comes into play. The emphasis is on methods that are reasonable to implement and appropriate for today's (and tomorrow's) data sizes, rather than on the asymptotics that one sees with the "theoretical approach."

Reading the book, I was taken with the thorough coverage of the topic and the clarity of presentation. Finding, easily, specific results was, well, easy, as suits the experienced researcher in the field. On the other hand, the careful exposition of key concepts, with elucidating examples, makes it ideal as a graduate text or for the researcher from a tangentially related area. The book covers the historical and mathematical background along with the key developments of the 1990s and early years of the current century, which form its core. Text indexing has been a major driving force for the area, and techniques for it are nicely covered. The final two chapters point to long-term challenges and recent advances. Updates to compact data structures have been a problem for as long as the topic has been studied. The treatment here is not only state of the art but will undoubtedly be a major influence on further improvements to dynamic structures, a key aspect of improving their applicability. The final chapter focuses on encodings, working with repetitive text, and issues of the memory hierarchy. The book will be a key reference and guiding light in the field for years to come.

J. Ian Munro
University of Waterloo