

Parsing Foundations

A Study of Data and Parsing

Unfinished Draft of First Edition

Parsing Foundations

A Study of Data and Parsing

Unfinished Draft of First Edition

Richard R. Masters
Seattle, WA



Self Publishers Worldwide
Seattle San Francisco New York
London Paris Rome Beijing Barcelona

This book was typeset using L^AT_EX software.

Preface

This book was created from a desire to make parsing easier to understand. The author struggled to learn a variety of topics in the field parsing. It was not that the ideas themselves were particularly difficult, but that it required a huge investment in both time and money to acquire, understand, and organize the relevant information. After all that effort, it was clear that parsing could be explained a lot better, and also that there is a lot of existing research, some of it several decades old, that could benefit from a wider audience.

Prior Work

Prior to the writing of this book, there were not that many resources for learning about parsing that were both freely available and of high-quality. There were no more than a handful of books about parsing on the market and the most relevant to a general audience was a relatively expensive book about compilers. There are also quite a few papers and articles written for academic journals which may not be accessible or understandable to a general audience. Finally, while there are some good online resources, they are lost within a sea of information that is either outdated or poorly produced, or only covers a limited set of topics.

In contrast, the goal of this book is to be accessible and as easy to understand as possible. This book is built from the ground up with relevant topics, starting from the basics and building up from there, while using a simple and intuitive style of presentation.

Some of the explanations in this book may seem excessively long so it is worth noting that keeping explanations short is a much lower priority for this book than providing a thorough explanation. The idea is that far more time is wasted by avoiding difficult material, skimming it repeatedly, and finally focusing your mind enough to understand it, than is wasted reading and understanding material that is excessively explained.

Audience

This book can serve as an introduction to topics on data and parsing for the working software developer or student of computer science. The book covers

foundational topics of parsing using concrete examples and common tools used in the field. In an academic setting, the book would serve well as a companion book for courses that only cover parsing as part of a larger topic, such as compiling.

Uses for this Book

The book may also be of interest to you if you are an educator, a blogger, or if you are writing a book. This document is licensed under the Creative Commons Zero v1.0 Universal license. You are free to copy this content, distribute it, include it in your own content, or sell it, and you may do so *without attribution*. Just copy what you want and move on. Read the license text at <https://creativecommons.org/publicdomain/zero/1.0/legalcode> for more details.

Prerequisites

For the chapters on data, the reader should have a basic knowledge of computers, such as knowing about a monitor, mouse, disk, file, and an application.

Table of Contents

1	Introduction
---	--------------

1

Chapter 1

Introduction

This book is about data and parsing in the context of computing. Data is defined by Wikipedia as any sequence of symbols for which "a specific act of interpretation" gives those symbols meaning. Parsing is an act of interpreting data based on how it matches structural patterns.

