Chapter 4: Data Processing

4.1 Introduction

Modern computers can process vast amounts of data representing many aspects of the world. From these big data sets, we can learn about human behavior in unprecedented ways: how language is used, what photos are taken, what topics are discussed, and how people engage with their surroundings. To process large data sets efficiently, programs are organized into pipelines of manipulations on sequential streams of data. In this chapter, we consider a suite of techniques process and manipulate sequential data streams efficiently.

In Chapter 2, we introduced a sequence interface, implemented in Python by built-in data types such as `list` and `range`. In this chapter, we extend the concept of sequential data to include collections that have unbounded or even infinite size. Two mathematical examples of infinite sequences are the positive integers and the Fibonacci numbers. Sequential data sets of unbounded length also appear in other computational domains. For instance, the sequence of telephone calls sent through a cell tower, the sequence of mouse movements made by a computer user, and the sequence of acceleration measurements from sensors on an aircraft all continue to grow as the world evolves.

Continue: 4.2 Implicit Sequences

Composing Programs by John DeNero, based on the textbook Structure and Interpretation of Computer Programs by Harold Abelson and Gerald Jay Sussman, is licensed under a Creative Commons Attribution-ShareAlike 3.0 Unported License.