

The tenth and final module of the book goes over the different types of data on the Internet and computational thinking. I have only thought of Internet data as files on a computer or server, but data could really be anything, such as facts on a website or weather information, but this module helped me understand that. I also never knew that there were many different ways to extract information from data. It goes to show that even though it may not seem like it, there are many aspects of data on the Internet and many ways to collect it.

The first part of the module is about data. Data is not only on the Internet, but it is everywhere. Anything with information can be considered data, like the scores of an exam. Data could also be derived from other data, such as the average of the exam scores. With this data, people could gain knowledge of the information. When data is organized into a collection, it is known as a dataset. An example of a dataset would be the inches of rain recorded on each day of the month. Valid data is data that has a meaning to it and represents the concept of it. In addition to data, there is also metadata, which is data about the data, such as where and when the data was collected. The use of data and its metadata leads to information, which is the truth that arises from the interpretation of data. Data could also be structured, or highly organized, into a spreadsheet, which is computer software used for storing and organizing data. With this software, data is organized into columns and rows that could be searched, sorted, modified, or deleted. A more advanced way of structuring data is to put the data in a database, a organized collection of tables, reports, and queries. When data has little to no organization, it is known as unstructured data. An example of unstructured data is a collection of pictures on a phone. Certain methods that data scientists also use relating to data include cleansing, which is finding and fixing corrupt data in a set, anonymization, which is removing personally identifiable information from a set, and aggregation, which is compiling information from multiple sets or databases into one single set or database.

The module then goes over computational thinking (CT), which is a method of problem solving that uses parts of both classic techniques and new computational techniques to make a highly effective framework. The main methods of computational thinking include abstraction, when essential parts of a problem are in a data model, decomposition, which partitions problems into certain components, generalization, when models are created from the analysis of limited data, pattern recognition, which uses machine learning to create predictive models, and algorithmic design, which creates a sequence of unambiguous steps that lead to a problem solution. There are also data structures, which are program-implemented constructs made to hold certain types of data. The most common structures include stacks, queues, linked lists, and binary trees. Stacks and queues are simple data structures that have a last in, first out storage model. Linked lists are sequential data structures which consist of nodes, which contains a data value and a pointer to the next node. A binary tree also uses nodes, but each node contains two pointers, one for the node on the left, and another for the node on the right. There is also a node at the top of a binary tree called a root node. Most of the time, in order to be a successful analyst, a more complex data structure will have to be used because datasets are much more efficient when they are sorted and could be searched multiple times for a value. The module then ends with going over interfaces and implementation. Application-programming interfaces, or API, is a set of protocols for writing programs that interact with the data in the application. Twitter, Flickr, and Facebook all use APIs in their webpages. Most programs have a user interface, which is how a human can interact with them. There is also implementation, which is a program that refers to how the steps of an algorithm are stored in a computer. User interface and implementation are usually used together, but the separation of them is one of the main principle of computational thinking and it is commonly used with ATMs, web browsers, and media players having a similar interface that is almost universal.