

Glossary

API

An Application Programming Interface (API) is a block of programming code that enables data to be transmitted between software, it provides a mechanism for two applications to talk to each other through the sharing of data.

Choropleth

A visualisation method using a map of a geographic area that uses differences in shading, colouring, or symbols within predefined areas to indicate the average values of a particular quantity in those areas, such as population, political affiliation, and household income.

Geo-location

Geographic location refers to a single position on Earth, and is defined by two coordinates, longitude and latitude, whereas Geolocation identifies the physical geographic location of objects, such as computers and mobile devices connected to the internet.

GeoJSON

An open standard format based on the JSON format and used to represent geographical features such as roads, and country, region, and city borders. As the name suggests.

Histogram

A graphical display of data using bars of different heights. More specifically a **histogram** represents the distribution of numerical values. Although they appear to be similar to a Bar Chart, a **histogram** is applied when we wish to group values into specific ranges. Consequently, the height of each bar in a histogram, actually shows how many of our data points fall into each of the ranges.

Heatmap

A visualisation method similar to a choropleth, the difference being that no pre-defined geographical areas are used for the visualisation.

High-level/low-level programming language

Computers fundamentally only understand a certain set of very basic instructions: store values, retrieve values, perform arithmetic operations, etc. In order for humans to program computers, it is convenient to use programming languages which allow us to describe what we would like the computer to do in more familiar terms, this language then gets translated into machine instructions.

Programming languages can be either relatively close to machine instructions (like 'assembler' languages), or be abstract but still expose a lot of the underlying machine to the programmer (like the languages C, D, or Rust), or



© Coventry University. Licensed under the Creative Commons Attribution-NonCommercial 4.0 International licence (CC BY-NC 4.0)

be abstract to the point where the underlying machine has little importance (like C#, Java, Go, or Python). We call these latter languages 'high-level' languages and the others 'low-level'. As a rule of thumb, high-level languages result in slower programs but are very user-friendly.

JSON

JavaScript Object Notation, or JSON for short, is a file and data format, that uses human-readable text to store and send data objects over a network or between software applications. JSON consists of key-value pairs (a key being equivalent to a field name in a database, and value the actual data value stored under that field name). JSON is a very common data format, used in a diverse range of applications, particularly when accessing API's, such as social media APIs.

Jupyter notebook

Jupyter (which combines the names of the three programming languages it supports: Julia, Python, and R) notebooks are a programming environment that runs on a client/server model: the client can ask the server to run program code and receives the "result" of running it. The usual client runs in the web-browser and follows the "notebook" format where we can alternative between source code and documenting text.

Library

A library in the context of programming is a collection of code that one can import in order to solve certain problems. For example, python provides the 'math' library with things like trigonometric functions or 'random' which enables one to generate random numbers. Libraries enable us to use code created by others so that we do not have to implement everything ourselves.

Machine learning

Machine learning is a collection of methods that "learn from data". Instead of hand-crafting a program that makes decision about data, we instead feed data into a machine learning program (alongside information about what it is supposed to "learn" from the data) and it will create something similar to a program (a "model") which can then make decisions based on the data.

Metric

A metric takes two mathematical objects (like vectors) and computes a distance between them. We use metrics as measures of similarity: the smaller the distance is, the more similar the objects are. Metrics are like distances in that they are always positive (no negative distances) and in that an object has distance zero to itself.

Perceptron

The **perceptron** is an algorithm for *supervised learning* of *binary classifiers*, that is whether an input (the data) falls into one of two categories, for example we may only be interested in whether the subject of an image can be classified as a cat or a dog. The perceptron algorithm classifies the input based on whether it, belongs to a specific class according to a set of features the developer engineers, and the resulting calculated weights after training.

Python

Python is a high-level programming language with a simple syntax and a large library of modules for e.g. math, scientific computation, language processing, web programming, and more.

Reverse geo-coding

Reverse geocoding describes the process of converting a geographical coordinate (latitude, longitude) to a readable address that can be used to extract the names of places, states, and countries.

Social media data

Data that is published, stored, and viewed through electronic communication, which includes social networking and microblogging websites, through which users form online communities to share information, ideas, personal messages, and other content, such as images and videos. Typical examples include Facebook, Twitter, Reddit, TikTok, and 4Chan.

Unparsable

When we work with large volumes of data it is likely that some of it will contain be corrupted and thus contain errors, or there may be human generated inconsistencies in the way the data has been partitioned that result in our code failing to process the data as you intended. Unparsable therefore refers to instances of data that have not been read correctly by your code, and can therefore not be stored or operated by the rest of your code later on.

Vector

In our context a vector is a sequence of numbers which we call 'entries'. The number of entries is called the dimension of the vector. Vectors can be used to represent coordinates; in that case we often combine vectors with a metric which defines a type of distance between them.