## Intro

The proliferation of computing has created a vast amount of data, there is data almost about everything. The use of these large data sets about things, such as logs of all visitors to a web site, provide opportunities for extracting information and knowledge but also creates new challenges as more data is coming in everyday. An example of collecting vast amounts of data to identify trends is when Google records all of their searches in a two day period which can lead to finding certain trending searches at certain times of day. These search tools are essential for finding information but the challenges it faces are filtering out data that is not of interest of the user searching for specific things. Information filtering systems take large data sets and eliminate this data that is not of interest. Software tools, including spreadsheets and databases help to efficiently organize and find trends in information.

## Body of the Article → Examples and Facts



<https://d1jnx9ba8s6j9r.cloudfront.net/blog/wp-content/uploads/2018/11/Untitled-1-768x183.jpg>

* **Web Scraping** is an example of one technique that uses computers to gather large amounts of information.
  + If someone wanted to know how many websites have the word “turtle” in their body it would take years to go through many websites and count the word turtle. Instead of doing this manually, we could create code to go through the websites for us and search for the specified word - in this case turtle - then get the count in a matter of minutes.
* **API Calls** are extremely important for allowing programs to communicate with each other which makes data gathering between programs possible.
  + Chances are whatever website you are using there are api calls in use to present the data you see.
* Computers are used in an iterative or interactive approach when processing digital information to gain insight and knowledge. Going through all data in large data sets and filtering and cleaning it is an iterative approach. Doing so by combining data sources, clustering data and classifying data are all parts of the process of using computers to process information.
* The interactive approach means that people can gain knowledge and insight when translating and transforming digitally represented information. Patterns can emerge that one might night have seen before when transformed using computational tools.

## Try it

* [Scratch: Dino Zoo Data Gathering](https://scratch.mit.edu/projects/360308465/editor)
* [Quick and Dirty Data Gathering with Python](https://towardsdatascience.com/quick-and-dirty-data-gathering-with-python-9d3d4b8cba13)
* [Web Scraping in Python](https://www.edureka.co/blog/web-scraping-with-python/)
* [Playing with Twitter Data Using Python](https://www.earthdatascience.org/courses/earth-analytics-python/using-apis-natural-language-processing-twitter/get-and-use-twitter-data-in-python/)

## Resources

* [What is an API?](https://www.youtube.com/watch?v=s7wmiS2mSXY)
* [Knowledge Base - 3.2 Large Data Sets](https://docs.google.com/document/d/1RS7vz3RL8zvucRUrypCCAzgB44qEfb3LdCpslcqmJ-c/edit?usp=sharing)

## Vocabulary

* Web Scraping
* API Call
* API