AIFE: UNIT-2
MS. SHWETA TIWARI
Published: APRIL, 2022

"Artificial Intelligence for Engineering/Engineers (KMC-201)"_

UNIT-2: DATA And ALGORITHMS

FALL SEMESTER, YEAR (2ND, 1ST)

FALL SESSION (2021-22)
(AIFE)
MS. SHWETA TIWARI
Published: APRIL, 2022

PREPARED FOR
Engineering Students
All Engineering College

PREPARED BY
SHWETA TIWARI
Guest Faculty

TOPIC ON: Everything about Data Processing | Definition, Methods, Types & Application

TOPIC ON: Everything about Data Processing | Definition, Methods, Types & Application

What do you mean by data processing?

Data processing is the conversion of data into usable and desired form. This conversion or "processing" is carried out using a predefined sequence of operations either manually or automatically. Most of the processing is done by using computers and thus done automatically. The output or "processed" data can be obtained in various forms. Examples of these forms include image, graph, table, vector file, audio, charts or any other desired format. The form obtained depends on the software or method of data processing used. When done itself it is referred to as automatic data processing.

Data processing is basically synchronizing all the data entered into the software in order to filter out the most useful information out of it. This is a very important task for any company as it helps them in extracting most relevant content for later use. Every important sector, be that banks, schools, colleges or big companies, almost all requires this processing of data. This processing is performed in order to store the most refined information in their systems for later use. Manual processing is very time consuming and requires you to engage too many people to do so. This is really not a feasible task when you have data in bulk. Nowadays industry people depend on strong and well efficient software tools to help in processing all that data. This helps them in achieving greater accuracy and enhance their efficiency. With the proper processing of data, more and more information can be sorted. This helps in getting a clearer view of matter and having a better understanding of it. This can lead to better productivity and more profits for the various business fields.

Real World Applications of Data Processing

With the implementation of proper security algorithms and protocols, it can be ensured that the inputs and the processed information is safe and stored securely without unauthorized access or changes. With properly processed data, researchers can write scholarly materials and use them for educational purposes. The same can be applied for evaluation of economic and such areas and factors. In the healthcare industry, the processed data can be used for quicker retrieval of information and even save lives. Apart from that, illness details and records of treatment techniques can make it less time-consuming for finding solutions and help in reducing the suffering of the patients.

Processing data to arrange it by type and information can save a lot of space taken up by data which is not organized and stored haphazardly. Processed data can also help in making sure that all staff and workers can understand it easily. They can implement it in the work, which can otherwise take up more time and end up in providing a decreased output. This can harm the interests of the business or organization.

Focus Of Data Processing

Most businesses and fields require data for providing a good quality of service. Having a collection of insights about collected data and their implications is a very important aspect of managing it and ensuring statistical authenticity. It is particularly essential for services concerned with financial technologies. This is so because transaction data and payment details need to be properly stored for easy access by customers as well as the company officials upon need. Processing is not limited to computers and can be done manually as well.

While the manual option uses brain power and intelligence, electronic data processing techniques can save a lot of time and ensure a smooth workflow and ensure adherence to deadlines. Accuracy is also higher with electronic processing. One of the essential aspects of this is to make sure that the insights formed are stored for future and shared use so as to save computational power and time.

Fundamentals of data processing & how data is processed

Processing of data is required by any activity which requires a collection of data. This data collected needs to be stored, sorted, processed, analyzed and presented. This complete process can be divided into 6 simple primary stages which are:

1. Data collection

2. Storage of data
3. Sorting of data
4. Processing of data
5. Data analysis
6. Data presentation and conclusions

Once the data is collected the need for data entry emerges for storage of data. Storage can be done in physical form by use of papers, in notebooks or in any other physical form. With the emergence and growing emphasis on Computer System, Big Data & Data Mining the data collection is large and a number of operations need to be performed for meaningful analysis and presentation, the data is stored in digital form. Having the raw data and processed data into digital form enables the user to perform a large number of operations in a short time and allows conversion into different types. The user can thus select the output which best suits the requirement.

This continuous use and processing of data follows a cycle called the data processing cycle and information processing cycle. These cycles might provide instant results or take time depending upon the need of processing data. Complexity in this field is increasing which is creating a need for advanced techniques.

Storage of data is followed by sorting and filtering. This stage is profoundly affected by the format in which data is stored. This further depends on the software used. General day and non- complex data can be stored as text files, tables or a combination of both in Microsoft Excel or similar software. As the task becomes complex which requires performing specific and specialized operations. They require different data processing tools and software which is meant to cater to the peculiar needs.

Storing, sorting, filtering and processing of data can be done by single software or a combination of software whichever feasible and required. Such a processing thus carried out by software is done as per the predefined set of operations. Most of the modern-day software allows users to perform different actions based on the analysis or study to be carried out. It provides the output file in various formats.

Different types of output files obtained as “processed” data

- **Plain text file** – These constitute the simplest form or processed data. Most of these files are user readable and easy to comprehend. Very negligible or no further processing is these types of files. These are exported as notepad or WordPad files.
- **Table/ spreadsheet** – This file format is most suitable for numeric data. Having digits in rows and columns allows the user to perform various operations. For ex, filtering & sorting in ascending/descending order to make it easy to understand and use. Various mathematical operations can be applied when using this file output.
- **Charts & Graphs** – Option to get the output in the form of charts and graphs is handy and now forms standard features in most of the software. This option is beneficial when dealing with numerical values reflecting trends and growth/decline. There are ample charts and graphs available to match diverse requirements. At times there exists a situation when there is a need to have a user-defined option. In case no inbuilt chart or graph is available then the option to create own charts, i.e., custom charts/graphs come handy.
- **Maps/Vector or image file** – When dealing with spatial data the option to export the processed data into maps, vector and image files is of great use. Having the information on maps is of particular use for urban planners who work on different types of maps. Image files are obtained when dealing with graphics and do not constitute any human readable input.
- **Other formats/ raw files** – These are the software specific file formats which can be used and processed by specialized software. These output files may not be a complete product and require further processing. Thus they will need to perform steps multiple times.

Methods of processing

1. **Manual Processing:** In this method data is processed manually without the use of a machine, tool or electronic device. Data is processed manually, and all the calculations and logical operations are performed manually on the data.
2. **Mechanical processing –** This is done by use of a mechanical device or very simple electronic devices like calculators and typewriters. When the need for processing is simple, this method can be adopted.

3. **Electronic processing** – This is the modern technique to process data. Electronic Data processing is the fastest and best available method with the highest reliability and accuracy. The technology used is the latest as this method used computers and was employed in most of the agencies. The use of software forms an integral part of this type. The data is processed through a computer; Data and a set of instructions are given to the computer as input, and the computer automatically processes the data according to the given set of instructions. The computer is also known as an electronic data processing machine.

Processing types on the basis of process/steps performed

There are various types of data processing, some of the most popular types are as follows:

1. Batch Processing
2. Real-time processing
3. Online Processing
4. Multiprocessing
5. Time-sharing

Why is it required?

- It is really difficult to work on raw data. Because every bit of information provided may not be that useful for you. You are required to filter out relevant content.
- You can't every time refer to that huge pile of raw data and select that relevant information you are looking for. This will make your work more tedious and bulky.
- Data processing will help you arrange the filtered out content into a homogenized form so that you can easily match those big figures as and when you require it.

- It will make it easy for you to look for any relevant information and also makes your work easy.
- It will even make this whole procedure more cost effective too. As arranging those big figures into well-structured tables saves you from that risk of losing your important information. And also some of the information gets filtered out thus the cost of saving that irrelevant information is also saved.
- It also makes it easier for you to modify and edit your processed data. You just have to look for similar cells and implement the same rule to all the cells you want to be modified.
- Data processing is very important before you start to do data mining. It reduces your cost of doing all the paperwork required to otherwise process the whole information and filter out all the relevant content manually.
- This increases the overall performance of any company as it rules out unnecessary steps that can hinder the whole data processing process.
- It automatically deletes all your duplicate documents and thus helps in making some storage space in your system.

What makes processing of data important

Nowadays more and more data is collected for academic, scientific research, private & personal use, institutional use, commercial use. This collected data needs to be stored, sorted, filtered, analyzed and presented and even requires data transfer for it to be of any use. This process can be simple or complex depending on the scale at which data collection is done and the complexity of the results which are required to be obtained. The time consumed in obtaining the desired result depends on the operations which need to be performed on the collected data and on the nature of the output file required to be obtained. This problem becomes starker when dealing with a very large volume of data. For example data collected by multinational companies. They collect data about their users, sales, manufacturing, etc. such services and companies dealing with personal information and other sensitive information must be careful about data protection.

The need for processing becomes more and more critical in such cases. In such cases, data mining and data management come into play without which optimal results cannot be obtained. Each stage starting from data collection to presentation has a direct effect on the output and usefulness of the processed data. Sharing the dataset with a third party must

be done carefully and as per written agreement & service agreement. This prevents data theft, misuse and loss of data.

What type of data needs to be processed

Data in any form and of any type requires processing most of the time. This data can be categorized as personal information, financial transactions, tax credits, banking details, computational data, images and simply almost anything you can think of. The quantum of processing required will depend on the specilisatized processing which the data requires. Subsequently it will depend on the output that you require. With the increase in demand and the requirement for such services, a competitive market for data services has emerged.

Important Data Processing Tools

1. **Surveying Tools** – SURVEY MONKEY, etc. software tools which help us in easily organizing those elaborated surveys to help us gather the relevant content from the right people.
2. **Statistical Tools** –SAS (STATISTICAL ANALYSIS SYSTEM) etc are statistical calculation tools that help in plotting those big graphs and charts to help us study certain relevant patterns and thus do effective comparisons and draw proper conclusions.
3. **Calculation and Analysis tools** – EXCEL and CALC, etc. are those mathematical software tools that help in applying relevant formulas to process the whole data.
4. **Database Management tools** – ACCESS and BASE, etc. are the tools that help us to manage a large amount of data that otherwise becomes too tedious to look after or refer to as and when we require to do so.