

Information Retrieval: Journey of research in relation to the multidimensional domain area of science

A. Murugaiyan

Introduction

Freedom of Expression happens when there is information, and information is nothing but Data or Datum in a computer field. The existing world uses this information at different levels with different technologies. Having data or information with no way to use them is like having a car with no driver. The technocrats build many systematic approaches to manipulate this data/information efficiently. Due to emerging technological influences and research there are ways to migrate from the base information into new information solutions. The field of computer science develops a strategical approach to magnify these technological influences, as a matter of research in data analysis. This yields mining, knowledge engineering, temporal data analysis, spatial data analysis, opinion data/information analysis, data optimization etc., as siblings of Information Retrieval. From the growth of internet technologies research in IR has started flourishing.

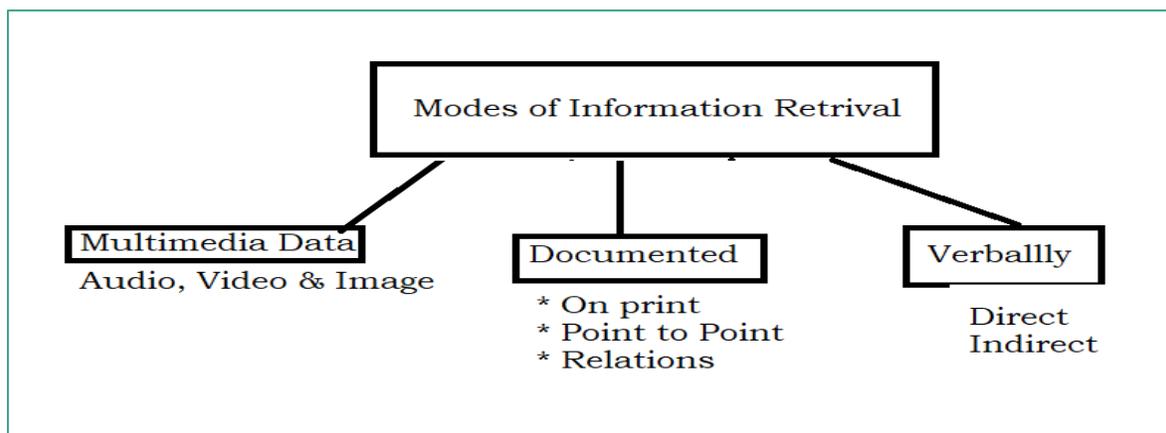


Figure 1. Architecture of IR

Information Retrieval is nothing but getting data/information from certain sources in general. Nowadays most information is stored in databases. Depending on the type of data, data can be classified as structural or nonstructural, and also according to its relationships within the information group; that is, the dependencies about the data mentioned in computer science. Information not only hides in databases, but is also available everywhere

in different formats: as social network data, sampling data, website data, etc. But it is not easy to compile or analyze such data, as their size and type are not compatible for human manipulation. Information Retrieval (IR) is the process which helps to overcome this difficulty.

If we retrieve such information as it is, it won't be much use due to the density of unnecessary data along with the useful data. To carry out the cleansing process, computer science provides a procedure called Data/Information Preprocessing. Data preprocessing is the first step for all IR methods. This preprocessing of data/information is done as follows:

1. Clean the inconsistent data
2. Integrate the data with different relations (identifying relationship patterns)
3. Data reduction

In this way we move forward in the research of information retrieval.

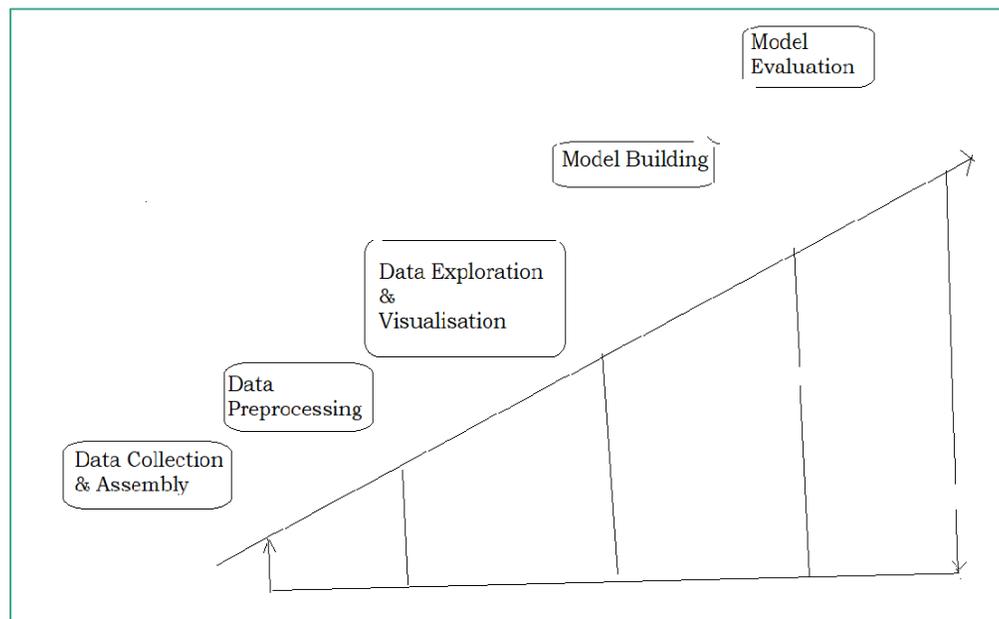


Figure 2. Stages of processing information

Method of research

Data Mining (knowledge discovery)

Data Mining is a process used to achieve useful information when working with a large pool of information/data. The place where such big data/information is stored is called a *Data Warehouse*. This is core for the Business Intelligence System which is built for data/information analysis.

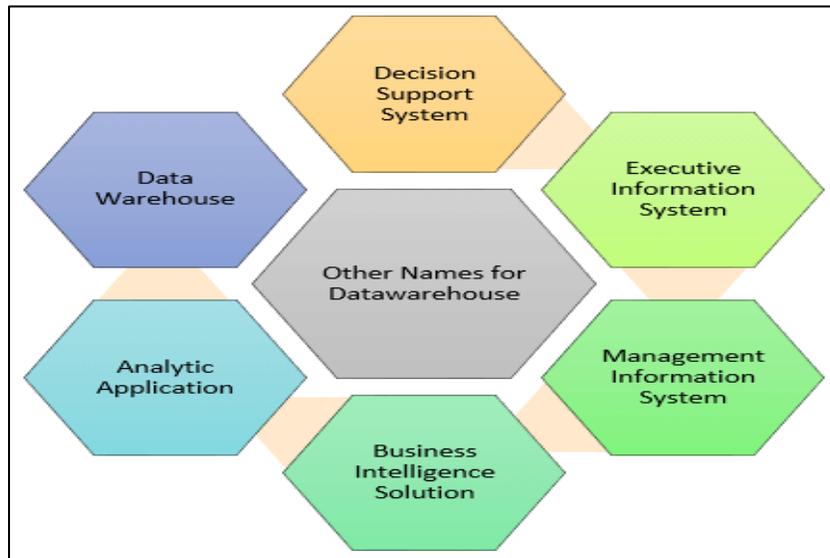


Figure 3. Names of IR

Page Ranking

We use browsers to search for information by typing text into the search bar. The relevant data will be found by the web crawlers, via search engines, and results are ranked based on number of visits with the help of a *Page Ranking Algorithm*. This is the gateway for the IR process in web-related information.

Temporal Mining:

Some data are available for a certain period of time in certain databases (Temporal Database). This information can be used within the period of availability.

Business Intelligence Process

Information retrieval is also a focus for business analytics. For example, the inventory data IR process which will reveal which products are selling fast, which type of products are purchased together; thus leading businesses to better procure their products or recommend their sales orders accordingly.

Metrological Data Analysis

Weather forecasting is another research sector in association with the IR process. Based on the previous data/ information, IR helps in analyzing the forecasted weather data.

Web Mining

This is a branch of research in Data Mining, which helps to retrieve hidden or useful information from web information/ data, for example *Heterogeneous Data* – a combination of several kinds of file formats. This allows the researchers to analyze all kinds of information such as structure, content and usage in web resources with the help of data analysis tools. In

other words, data mining in web data/information is called *Web Mining*. While analyzing heterogeneous data, researchers are influenced by some types of mining such as Text Mining, Image Mining, Opinion Mining, etc.

The availability of different options for research proves that no data is useless data; it is just which data is useful data, which is determined through the Information Retrieval Process.

Text Mining

It is branch of research in data mining which analyzes the content of the web page/list of textual information. Through the preprocessing techniques, it removes duplication and makes sense relevant to the text, and also finds patterns or any relevant hidden solutions.

Opinion Mining

Finding solutions related to opinion information which is posted in social media is *Opinion Mining*. People post/share their feelings on social networks through images and text. IR provides techniques to mine this information and recommend to the people any information based on their interests and their moods. Following this research IR assists with image processing.

Knowledge Discovery

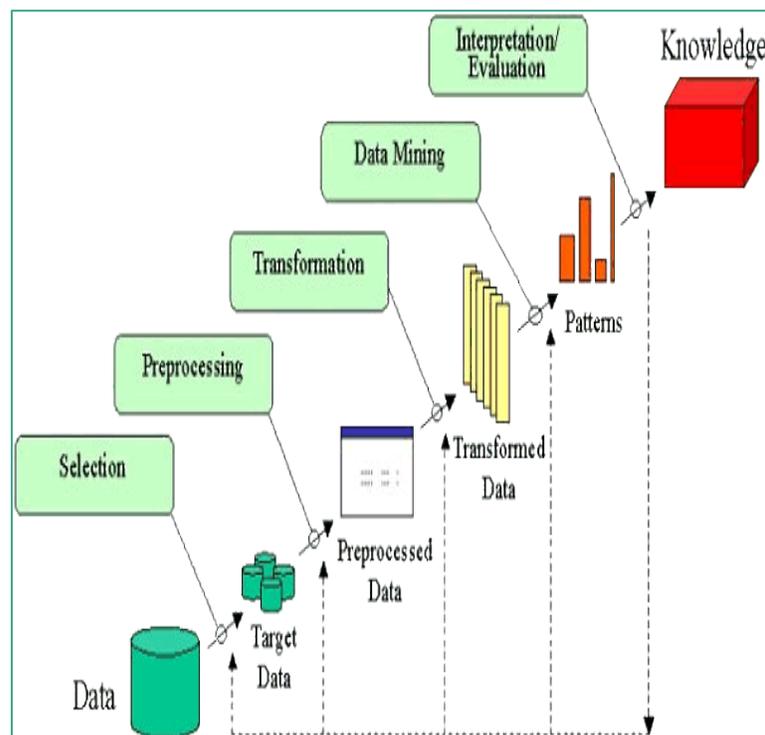


Figure 4. The process of Knowledge Discovery database

The final target of the IR process is *Knowledge Discovery*. It is the application of techniques using data mining, machine learning, pattern recognition, statistics, artificial intelligence, until identified knowledge is reached. Each technique is a pathway of research in different domains since computer science is combination of logical and mathematical science. Nowadays this research approach is implemented in all disciplines.

IR in interdisciplinary domains

The existing research requires statistics to show all the analytical results. In order to achieve this, Information Retrieval plays vital role as it contains all kinds of information and essential tools. When one needs any information, the probability of getting a solution has become at least more than one. (eg.: google search) and thanks to IR.

The IR system started, in 1945, by manipulating library data through manual processes. The first computer-based searching was initiated in 1948. Since 1970 different techniques have been developed for small levels of data; and now it has reached almost every technology, with different milestones.

To give an example, it has reached the level of producing efficient results in the field of medical science, even in research into genomes (analysis of genetical data). And, closer to home, being an island where conservation and preservation is very important, Seychelles can use mining techniques related to environmental science, as we have large data to be analyzed using ocean mining, pollution mining, etc.

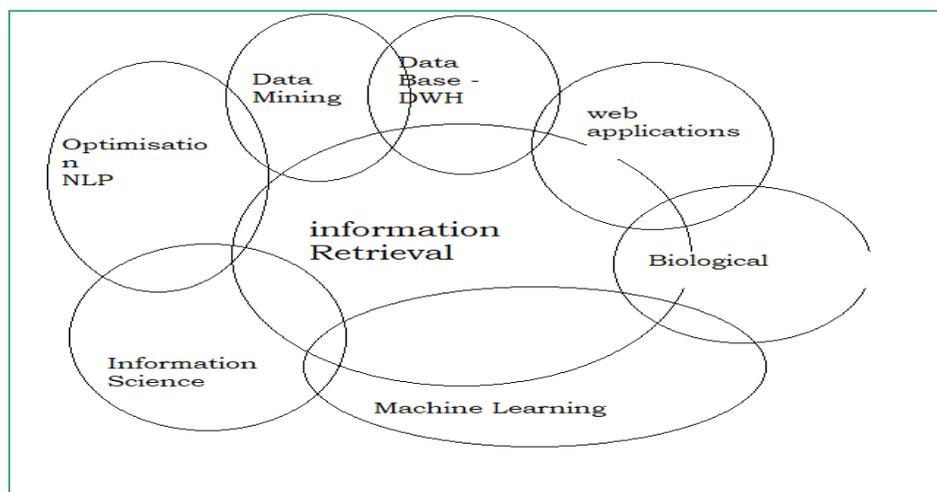


Figure 5. Roles of IR in different areas of research

Conclusion

IR is a process of identifying information from any storage medium with a systematic algorithm for different domain specifications. However, data/information is nowadays available in digital format. Research related to data analysis is more popular irrespective of the research domain. The study of information retrieval is not just related to computer science anymore; it is a combination of mathematics and logical science. Without mathematical support, researchers would not be able to prove their results; IR facilitates this to enrich the result of data analysis. The areas of research in IR are domain-interdependent, wherever information is available.

References

- Blummer, B. et al. (2014). *Information Research and Search Process*. Science Direct.com
Pei, J. et al. (2016). *Data Mining Concepts* (3rd ed.). Elsevier.
Tober, M. et al. (2014). *SEO Ranking Factors & Rank. Generation Google US*. yoonweb.com
Yiling, L. et al. (2018). Test case prioritization in regression testing. *Information & Software Technology*. 93 (C).

Alexander Murugaiyan is an academic computer professional and is currently working as a Senior Lecturer/Lecturer in the Computer Science Department at the University of Seychelles. He completed his Masters in Computer Application, and Master of Philosophy in Computer Science, at Bharadhidhasan University, India. His research area is Information Retrieval, Data Mining Techniques, Software Engineering and Project Management.