*RIK International Journal of Science and Technological Research (RIK-IJSTR)* *Vol. 7 No. 1, January 2021*
*ISSN: 2350-6310 (Print)*
*ISSN: 2350-543X (Online)*

# Information Retrieval and the Relational: An Enhanced Overview

**Humphery Uche**
Research Scholar
Department of Computer Science
Ignatius Ajuru University of Education
Rivers State
Nigeria

**&**

**Asagba, P. O.**
Department of Computer Science
University of Port-Harcourt
Rivers State
Nigeria

***ABSTRACT***

***In this paper information Retrieval and Relational Model were discussed and analysed. The former deals with the structure, analysis, organization, storage, searching, and retrieval of information in an unstructured document to satisfy a user's information needs while later are concerned with structured data retrieval through well-defined formal languages for representation and manipulation based on the theoretically founded data models. Within the computer science discipline, Relational Model (RM) and Information Retrieval (IR) systems are closely related fields. Information Retrieval centers on unstructured search with possibly vague query or search semantics and without a well-defined logical schematic representation while deals with structured data. Structured data refers to any data that resides in a fixed field within a record or file. This includes data contained in relational databases and spread sheets. By unstructured information, we generally mean information that does not have a well-defined formal model and corresponding formal language for representation and reasoning, but rather is based on understanding of natural language. This document compares these two related field to discover their difference and similarities.***

***Keywords: Database, Information Retrieval, Model.***

## Introduction

Information Systems are implemented within an organization for the purpose of improving the effectiveness and efficiency of that organization. Two paradigms that characterize much of the research in the Information System discipline are behavioral science and design science. The behavioral science paradigm seeks to develop and verify

theories that explain or predict human or organizational behavior. The design science paradigm seeks to extend the boundaries of human and organizational capabilities by creating new and innovative artifacts. The design science paradigm has its roots in engineering and the science of the artificial. It is fundamentally a problem solving paradigm and seeks to create innovations that define the ideas, practices, technical capabilities, and products through which the analysis, design, implementation, management, and use of information systems can be effectively and efficiently accomplished.

Usability is one of the focuses of Computer-Human Interaction. As the name suggests, usability has to do with bridging the gap between people and machines. For a tool to be effective, it must allow intended users to accomplish their tasks in the best way possible. Usability depends on a number of factors including how well the functionality users' needs, how well the through the application. The key principle for maximizing usability is to employ an iterative design, which progressively defines the design through evaluation from the early stages of design. The evaluation steps enable the designers and developers to incorporate users' and clients' feedback until the system reaches an acceptable level of usability. The preferred method for ensuring usability is to test actual users on a working system [1]

**Overview of Database Management System History**

It has been already seen that the ancestor to the DBMS was the file-based system. However, there was never a time when the database approach began and the file based system stopped. In fact, the file-based system still exists in particular areas. It has been suggested that the DBMS had its roots in the 1960s Apollo moon-landing project, which was begun in response to President Kennedy's objective of landing a man on the Moon by the end of that decade. At that time, there was no system available that would be able to handle and manage the massive amounts of information that the project would create Consequently, North American Aviation, the prime contractor for the project, developed software known as Generalized Update Access Method [2]. Generalized Update Access Method was based on the concept that smaller components come together as parts of larger components and so on, until the last is collected. This structure, which adapts to an upturned tree, is also known as a hierarchical structure. In the mid-1960s, IBM combined NAA to develop Generalized Update Access Method into what is now known as Information Management System [3]

The cause that IBM restricted IMS to the management of hierarchies of records was to allow the use of serial storage devices, most especially magnetic tape, which was a market need at that time. This restriction was later dropped. One of the earliest commercial DBMSs, IMS is still the main hierarchical. In the mid- 1960s, another important development was the appearance of IDS (or Integrated Data Store) from General Electric. This work was supervised by one of the early pioneers of database system, Charles Bachmann. This development led to a new type of database system known as the network DBMS, which had a deep effect on the information system of that

generation. The network database was developed partly to meet the need to represent more complex data relationships than could be modelled with hierarchical structures and partly to enforce a database standard [4]

To help set up such standards, the Conference on Data Systems Languages, comprising representatives of the U.S. government and the world of business and commerce, formed a List Processing Task Force in 1965, later renamed the Data Base Task Group in 1967. The terms of references for the system to define standard specifications for an environment that would allow database creation and data manipulation. The 1990s saw the growth of the internet, the three-tier client-server architecture. The late 1990s saw the development of XML (eXtensible Markup Language), which has had a deep effect on many aspects of IT, including database integration, graphical interfaces, embedded systems, distributed system and the database systems.

**Elements of a Database Management System (DBMS)**

It has been identified in the DBMS environment five major elements: people, procedures, data, software and hardware. The first component is the people involved with the system. There are four different types of people who are involved in the environment of a DBMS, application developers, database administrators, and end-user and database designers.

This component refers to the rules and instructions which manages the design and use of the database. The staff and the user of the system who run the database need to document the procedures on how to use or operation of the system. Maybe the most significant part of the DBMS environment is the data, where data is a collection of facts stored in the database Software element consists of database management system software and application software, in addition to the operating system and contains software network in the case that the DBMS used through the network. The applications and the DBMS need the hardware to run. Hardware can range from a single personal computer to a single central or the network of computers. Hardware depends on the specific requirements of the organization and database management systems are used. In some DBMS run only on specific operating systems or hardware, while others run on a variety of operating systems and hardware. This development led to a new type of database system known as the network DBMS, which had a deep effect on the information system of that generation [5].

**Advantages/Disadvantages of a DBMS Model**

The DBMS serves as an intermediary between the user and the database. The structure of the database itself is stored all application requests and translate them into the complex processes required to meet as a group of files and the only way to reach the data in those files through database management systems. The DBMS hides much of the database's complexity from the application programs and users. Although the database management system is characterized by a number of properties, unfortunately it has several disadvantages. In the following, advantages and disadvantages are listed [6].

**i) Advantages:**
a) Control the redundancy of the data.
b) Stability of the data.
c) Learn more about the same amount of data.
d) Data exchange.
e) Improving the integrity of the data.
f) Improving the security.
g) Implementation of standards.
h) Scale economies.
i) Improving access and response to the data.
j) Increasing end-user productivity.
k) Improving the maintenance through the independence of the data.
l) Increased synchronization.
m) Improving the services of backup and recovery.

**ii) Disadvantages:**
a) Difficulty.
b) Size.
c) The cost of a DBMS.
d) The cost of additional hardware.
e) The cost of conversion.
f) Performance.

**Operations in DBMS Relational Model**

Four basic update operations performed on relational database model are Insert, update, delete and select.

a) Insert is used to insert data into the relation
b) Delete is used to delete tuples from the table.
c) Modify allows you to change the values of some attributes in existing tuples.
d) Select allows you to choose a specific range of data.

Whenever one of these operations is applied, integrity constraints specified on the relational database schema must never be violated. Best Practices for creating a Relational Model

e) Data need to be represented as a collection of relations
f) Each relation should be depicted clearly in the table
g) Rows should contain data about instances of an entity
h) Columns must contain data about attributes of the entity
i) Cells of the table should hold a single value
j) Each column should be given a unique name
k) No two rows can be identical
l) The values of an attribute should be from the same domain

**Advantages of using a DBMS Relational model**

**a) Simplicity:** A relational data model is simpler than the hierarchical and network model.

**b) Structural Independence:** The relational database is only concerned with data and not with a structure. This can improve the performance of the model.

**c) Easy to use:** The relational model is easy as tables consisting of rows and columns is quite natural and simple to understand

**d) Query capability:** It makes possible for a high-level query language like SQL to avoid complex database navigation.

**d) Data independence:** The structure of a database can be changed without having to change any application.

**e) Scalable:** Regarding a number of records, or rows, and the number of fields, a database should be enlarged to enhance its usability.

**Disadvantages of using Relational model**

a) Few relational databases have limits on field lengths which can't be exceeded.

b) Relational databases can sometimes become complex as the amount of data grows, and the relations between pieces of data become more complicated.

c) Complex relational database systems may lead to isolated databases where the information cannot be shared from one system to another.

**The Concept of Information Retrieval**

Historically, information retrieval is "the discipline that deals with the structure, analysis, organization, storage, searching, and retrieval of information" as defined by Gerald Salton, an IR pioneer [7]. We can enhance the definition slightly to say that it applies in the context of unstructured documents to satisfy a user's information needs. This field has existed even longer than the database field, and was originally concerned with retrieval of catalogued information in libraries based on titles, authors, topics, and keywords.

In academic programs, the field of IR has long been a part of Library and Information Science programs. Information in the context of IR does not require machine-understandable structures, such as in relational database systems. Examples of such information include written texts, abstracts, documents, books, Web pages, e-mails, instant messages, and collections from digital libraries. Therefore, all loosely represented (unstructured) or semi-structured information is also part of the IR discipline.

**Information Retrieval (IR) Model**

Mathematically, models are used in many scientific areas having objective to understand some phenomenon in the real world. A model of information retrieval predicts and explains what a user will find in relevance to the given query. IR model is basically a pattern that defines the above-mentioned aspects of retrieval procedure and consists of the following:

a) A model for documents.
b) A model for queries.
c) A matching function that compares queries to documents.

Mathematically, a retrieval model consists of

D − Representation for documents
R − Representation for queries
F − The modelling framework for D, Q along with relationship between them.
R (q,di) − A similarity function which orders the documents with respect to the query. It is also called ranking.

**Types of Information Retrieval (IR) Model**

An Information Retrieval Model (IRM) can be classified into the following three models:

**a) Classical IR Model:**

It is the simplest and easy to implement IR model. This model is based on mathematical knowledge that was easily recognized and understood as well. Boolean, Vector and Probabilistic are the three classical IR models.

**b) Non-Classical IR Model:**

It is completely opposite to classical IR model. Such kinds of IR models are based on principles other than similarity, probability, Boolean operations. Information logic model, situation theory model and interaction models are the examples of non-classical IR model.

**c) Alternative IR Model:**

It is the enhancement of classical IR model making use of some specific techniques from some other fields. Cluster model, fuzzy model and latent semantic indexing (LSI) models are the example of alternative IR model.

**d) The Boolean Model:**

It is the oldest Information Retrieval (IR) model. The model is based on set theory and the Boolean algebra, where documents are sets of terms and queries are Boolean expressions on terms. The Boolean model can be defined as −

a) D − A set of words, i.e., the indexing terms present in a document. Here, each term is either present (1) or absent (0).
b) Q − A Boolean expression, where terms are the index terms and operators are logical products − AND, logical sum − OR and logical difference − NOT
c) F − Boolean algebra over sets of terms as well as over sets of documents
   If we talk about the relevance feedback, then in Boolean IR model the Relevance prediction can be defined as follows −
d) R − A document is predicted as relevant to the query expression if and only if it satisfies the query expression as –

e) Boolean logic is an essential tool in information retrieval and allows you to combine search terms. When you need more than one word to describe your search problem, you can combine multiple search terms with Boolean operators. The most common operators are AND, OR and NOT.
f) The AND operator is used to retrieve results that contain the entire search terms used.

**Summary and Conclusion**

It has been identified in the DBMS environment five major elements: people, procedures, data, software and hardware. The first component is the people involved with the system. There are four different types of people who are involved in the environment of a DBMS, application developers, database administrators, and end-user and database designers. This component refers to the rules and instructions which manages the design and use of the database. The staff and the user of the system who run the database need to document the procedures on how to use or operation of the system. Maybe the most significant part of the DBMS environment is the data, where data is a collection of facts stored in the database.

Software element consists of database management system software and application software, in addition to the operating system and contains software network in the case that the DBMS used through the network. The applications and the DBMS need the hardware to run. Hardware can range from a single personal computer to a single central or the network of computers. Hardware depends on the specific requirements of the organization and database management systems are used. In some DBMS run only on specific operating systems or hardware, while others run on a variety of operating systems and hardware. This development led to a new type of database system known as the network DBMS, which had a deep effect on the information system of that generation.

## References

[1] Agunna M. (2011). Info. Data challenge: A Data Management Perspective, *International Journal of Computer Application, Research and Software Engineering (IJCARSE) 2(1), 44 – 59.*

[2] Asuquo N. (2016). Comparative study of the new generation agile, scalable, high performance Relational Database Management System, *International Journal of Computer Application (IJCA), 48(20), 14 – 21*.

[3] Brekori N. (2013). Type of Databases and its role in the management of Info. Data. *International Journal of Applied Information Systems (IJAIS), 5(4), 2249 – 2266.*

[4] Charity R. (2014). Introduction to Info, Data, *An International Article for Swedish Institute of Computer Science.*

[5] Ezekoli S. (2012). Types of SQL Database and its comparison with Info, Data Management, *International Journal of Applied Information Systems (IJAIS) 5(4), 12 – 19.*

[6] Femi Z. (2012), Exploiting context for combining multiple entity resolution systems. *An International Journal of Applied Information System (IJAIS) 3(2), 10 – 15.*

[7] Greggory .J. (2012), Comparative study of the new generation agile, scalable, high performance Relational Database Management System, *International Journal of Computer Application (IJCA), 48(20), 14 – 21.*