

Architectural Framework for Development of Fuzzy Integrated Data Repository (FIDR) for Medicinal Plants in Nigeria.

Igbape, Emadomi Moses (PhD)¹, Enoyoze, Goddidit Esiro (PhD)²

¹Department of Computer Science, Auchi Polytechnic, Auchi, Nigeria.

²Department of Biological Sciences (Plant Biology and Biotechnology), Edo State University Uzairue, Nigeria.

DOI: <https://doi.org/10.51584/IJRIAS.2024.90216>

Received: 17 January 2024; Accepted: 27 January 2024; Published: 11 March 2024

ABSTRACT

Nigeria is a tropical country known to be home to a vast array of plants, most of which are plants that have been found to be medicinal and have been used in traditional healthcare delivery for centuries. Traditional Medicine Practitioners (TMPs) have been the major custodians of the knowledge of herbal medicine in Nigeria. It is noted that there is lack of documentation of the practice of herbal medicine which has resulted in the loss of a vast majority of the knowledge. The absence of a centralized digital data repository on medicinal plants and their uses in Nigeria hampers the drive for standardization, scientific validation and quality control of herbal medicine products in Nigeria. This consequently undermines the health and wellbeing of those dependent on traditional herbal medicine for their primary health care needs. Much of the claims on the efficacy of herbal medicine are found to be based on traditional beliefs and myth and have not been subjected to scientific evaluation. This accounts for the introduction of fuzzy logic into the design of an architectural framework for design of Fuzzy Integrated Data Repository that will allow access to information using sharp and fuzzy queries. This will be a resource platform that will resolve the information impasse by providing adequate and correct information to all stakeholders in the traditional herbal medicine healthcare delivery as well as foster synergy between orthodox medicine and herbal medicine practitioners, improve research that will lead to standardization and quality control of herbal medicine products in Nigeria.

Keywords: Medicinal plants, Herbal medicine, Fuzzy integrated data repository, Traditional Healthcare delivery

INTRODUCTION

Nigeria is a tropical country in the west coast of Africa with abundance of rain forest having a vast array of plants diversity. Most of the plants in Nigeria have been found to be medicinal and have been used in traditional healthcare delivery for centuries. According to the report by National Institute for Pharmaceutical Research and Development (NIPRD), over 60 percent of the Nigerian population depend on herbal remedies for their primary healthcare needs.

The practice of herbal medicine in Nigeria is mostly unique to the various indigenous communities and cultures. Over the years, Traditional Medicine Practitioners (TMPs) have been the major custodians of the knowledge of herbal medicine in Nigeria. The TMPs have also been the facilitators of herbal medicine practice in Nigeria. The knowledge and practices are usually handed down from generation to generation through oral communication as family heritage and are mostly shrouded in mystic. Lack of documentation

of this indigenous knowledge has resulted in the loss of a vast majority of indigenous information.

Igbape and Enyoze (2019) made a call for improved methods, principles and practices of documenting the information on African traditional medicines with the use of Information Technology tools in order to develop efficient herbal medicine data repositories.

There is currently a crunch in centralized and accessible resource for TMPs, researchers, and the general public to access information on herbal medicine practices in Nigeria. The observed dearth of documented information that can easily be accessed presents a number of issues;

(1) It is difficult to aggregate and preserve the traditional herbal medicine knowledge which is key to Nigeria's traditional medicine heritage.

(2) The absence of a centralized digital data repository on medicinal plants and their uses in Nigeria hampers the drive for standardization, scientific validation and quality control of herbal medicine products in Nigeria. This consequently undermines the health and wellbeing of those dependent on traditional herbal medicine for their primary health care needs

(3) The lack of a comprehensive digital repository of herbal medicine practice in Nigeria constitutes a drawback on research efforts aimed at identifying bioactive compounds and understanding of the mechanisms of actions of medicinal plants used in Nigeria. Thus a synergy between orthodox medicine practitioners and TMPs has become difficult to achieve.

The main aim of this work is to provide an Architectural Framework for the development of a Fuzzy Integrated Data Repository (FIDR) on Nigerian medicinal plants.

The objectives of this work are as follows:

1. To carry out indebt investigation on medicinal plants in Nigeria in order to be able to classify them as entities with their attributes. This will enable the creation of dimension tables and facts tables in the FIDR.
2. To provide architecture for the process of classification of medicinal plants based on their attributes using Fuzzy Classification Technique.
3. Provide architectural framework for development of FIDR with attributes that can be used to query the repository using both classical and fuzzy (plain linguistic terms) queries.

This research work seeks to proffer answers to the following questions:

1. What architectural framework can be used to develop herbal medicine data repository that will improve patients' outcomes and facilitate equitable delivery of healthcare services in Nigeria especially to underserved and remote communities?
2. What are the key architectural components that are critical for the development FIDR for medicinal plants in Nigeria?
3. How can the integration of Fuzzy logic into the architectural framework for the design of FIDR for medicinal plants in Nigeria be achieved?

Significantly, the creation of FIDR for herbal medicine in Nigeria will play a major role as follows:

1. Preservation of Nigerian herbal medicine heritage.
2. Conservation and sustainable use of medicinal plants in Nigeria.
3. Creating a data bank and a reference resource for the vast collection of data on medicinal plants in Nigeria.

4. It will provide relevant information for research to identify new compounds and development of new drugs.
5. It will help to provide information that will facilitate evaluation of the safety and efficacy of traditional herbal medicine.

THEORETICAL FOUNDATION

Digital herbal medicine information and practices in Nigeria requires that each medicinal plant be identified as an entity with specific attributes to be stored in a database that would be available for query purposes. It is observed that most attributes of medicinal plants are not precise. For instance, uses of herb, dosage, and efficacy etc. are claims based on human reasoning and are subjective. A particular herb might even be used for treatment of a variety of ailments. Due to the imprecise and vague nature of most of the attributes to be used for classification and analysis of medicinal plants, fuzzy logic is introduced to analyze and evaluate the fuzziness, vagueness or imprecision in the attributes of each medicinal plant.

The theory of fuzzy sets was introduced by Lofti A. Zadeh (1965) and expanded considerably in the works of Prade (1984) as well as Zimmermann (1991). The original interpretation of fuzzy sets arises from a generalization of the classical concept of subset extended to embrace the description of “vague” and “imprecise” notions. The generalization is made considering that the membership of an element to a set becomes “fuzzy” or “vague” concept. In the case of some elements, it may not be clear if they belong to a set or not. As in the case of classifying a herb based on its assumed efficacy for the treatment of an ailment. Then their membership may be measured by a degree commonly known as the “membership degree” of that element to the set and it takes a value in the interval (0,1) by agreement (Galindo, 2008).

With classical logic, it is possible to handle information that is completely true or completely false. It is not possible to handle information that is imprecise, vague or incomplete. However, in real life (especially in the domain of African traditional herbal medicine, where claims are based on traditions and myth), this can be the sort of information that is available.

Fuzzy logic is therefore an extension of classical logic (Zadeh, 1992). It deals with approximate reasoning instead of exact reasoning (Galindo, 2008). The importance of fuzzy logic lies in the fact that human reasoning, especially reasoning based on common sense and oral traditions are by nature imprecise.

By definition, a fuzzy set A (Zadeh, 1965) over a universe of discourse X (a finite or infinite interval within which the fuzzy set can take a value) is a set of pair

$$A = \{ \mu_A(x)/x : x \in X, \mu_A(x) \in [0,1] \in R \}$$

Where $\mu_A(x)$ is called a membership degree of the element x to the fuzzy set A . This degree ranges from between the extremes 0 and 1. Often the dominion of the real numbers $\mu_A(x)=0$ indicates that x in no way belongs to the fuzzy set A , while $\mu_A(x)=1$ indicates that x completely belongs to the fuzzy set A . $\mu_A(x)=0.5$ indicates the greatest uncertainty point.

Fuzzy classification is a machine learning approach which can be used to classify medicinal plants based on several attributes and data sets. Such as chemical constituents and therapeutic properties and can be used to develop efficient and reliable data repository.

In a study by Jin *et al.* (2012), the herbal properties and the biological performance characteristics of Chinese material medica with an either cold or hot herbal property were ascertained by using the fuzzy evaluation model.

In a study by Fasel and Shahzad (2012), they presented an approach to achieve the requirements of realizing Fuzzy Integrated Data Repository (FIDR) by using Fuzzy Dimensional tables for classification by mapping crisp dimensional tables in a 1:1 relationship. With this approach, an attribute to be handled fuzzily is extended with two meta tables; the first meta table contains the description of the fuzzy concept, the second meta table contains the membership degree of each instance with regard to the membership classes.

This approach has the advantage that numeric (crisp) values and fuzzy classification using linguistic variables can be accessible to users so the resulting database can be analyzed sharp and fuzzy.

ARCHITECTURAL FRAMEWORK DESIGN METHODOLOGY

The methodology of this research is a fusion of qualitative and quantitative approach. It adopts design science approach which identifies the problem and motivation in the domain of interest as a global view. It sets objects for the intended solution and resolves the design into specific solution components until specific solutions are modeled. Consequently, the study focused on extensive review of available literature on the domain of interest on modeling of FIDR so as to be able to follow established engineering procedures to model an FIDR that could resolve the problem of inappropriate documentation of medicinal Plants information in Nigeria or complete lack of it. The Kimball (2002) model of data warehouse shown in figure 2 is adopted with the components.

From the preliminary investigations in the process of this research work, it has been ascertained that huge volumes of data are encountered in the practice of herbal medicine in Nigeria. These data occur in the form of numeric data, text data, images, video, sound formats, geographic information (such as natural habitat, regions of cultivation, climatic conditions for cultivation) etc, with varying degrees of evidences of incompleteness, vagueness, imprecision and subjectivity. Aggregating these data for the purpose of information generation is cumbersome.

There are four separate and distinct FIDR components to be considered in an integrated data repository environment figure 1 (Kinball, 2002). These are:

1. Operational Source Systems (Data Sources)
2. Data Staging Area
3. Data presentation Area
4. Data Access Tools

OPERATIONAL SOURCE SYSTEMS (Data Sources)

This denotes the independent Sub systems and processes with which data are acquired for the FIDR such as surveys, interviews, field trips, questionnaires, observations etc. which can be viewed as transaction processing systems. The source data come from these processing systems. The main thrust of this work is to integrate data from these diverse independent sources and applications into a single intelligent system that can provide analysis and reporting functionalities of the FIDR.

DATA STAGING AREA

The data staging area is part of the FIDR and it is both a storage area and a set of processes commonly referred to as Extract – Transformation – Load (ETL). The data staging area is everything between the operational source systems and the data presentation area. The key architectural requirement of the data staging area is that it is off – limits to business users and does not provide query and presentation service. Once the data is extracted into the staging area, there are numerous potential transformation such as

cleansing the data, (correcting misspellings resolving domain conflicts, dealing with missing elements, or parsing into standard formats), combining data from multiple sources, reduplicating data and assigning repository keys. These transformations are all precursors to loading the data into the FIDR.

Another architectural requirement of the FIDR is that the presentation area should be dimensionally structured. The final step of the ETL process is the loading of data into the presentation area. This usually takes the form of presenting the quality – assured dimensional table to the bulk loading facilities of each data mart. The target data mart must then index the newly arrived data for query performance.

DATA ACCESS TOOLS

The final major component of the FIDR environment is the data access tool(s). This refers to the variety of capabilities that can be provided to TMPs and other stakeholders to leverage the presentation area for analytical decision making. By definition, all data access tools query the data in the FIDR’s presentation area, which is obviously the whole essence of using the FIDR.

A data access tool can be as simple as an ad-hoc query tool or as complex as a sophisticated data mining or modeling application (Kimball, 2002).

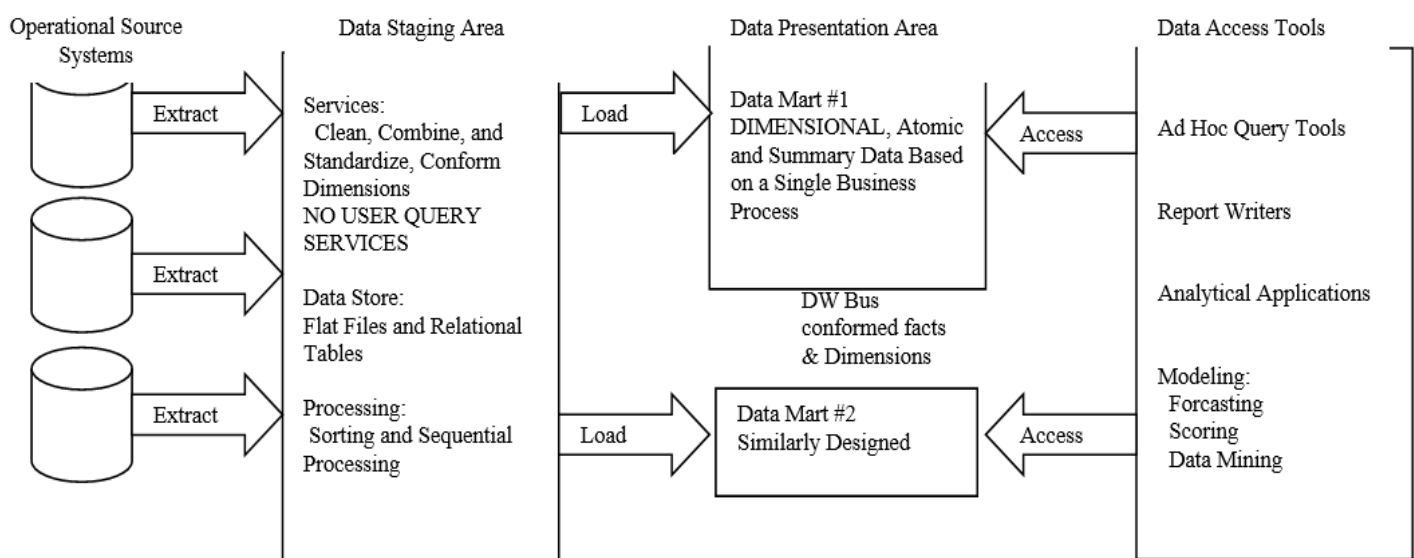


Fig. 1: Basic Elements of the Data Warehouse (Kimball, 2002)

Thereafter the research work provides an extension to Kimball’s model by integrating fuzzy concepts into the model as shown in figure 2. This is to resolve problems of imprecision, vagueness and fuzziness in the data encountered in documentation of herbal medicine practice as shown in figure 2.

Finally the architecture of the FIDR is modeled with the detail of the data transformation processes required as shown in figure 3 and figure 4 respectively. The modelling concepts are shown in details with implementation framework shown in figure 4.

The model presents the architecture in four main components as follows:

- Source data identification, collection and storage – operational source systems and database layer.
- Data analysis and transformation – data staging layer.
- Data presentation and reporting – data presentation layer and
- System implementation layer

ARCHITECTURE FOR THE FIDR

Figure 2 below presents a schema of the proposed FIDR model. This section also defines the fundamental concepts, which are building blocks used in the architecture of the FIDR.

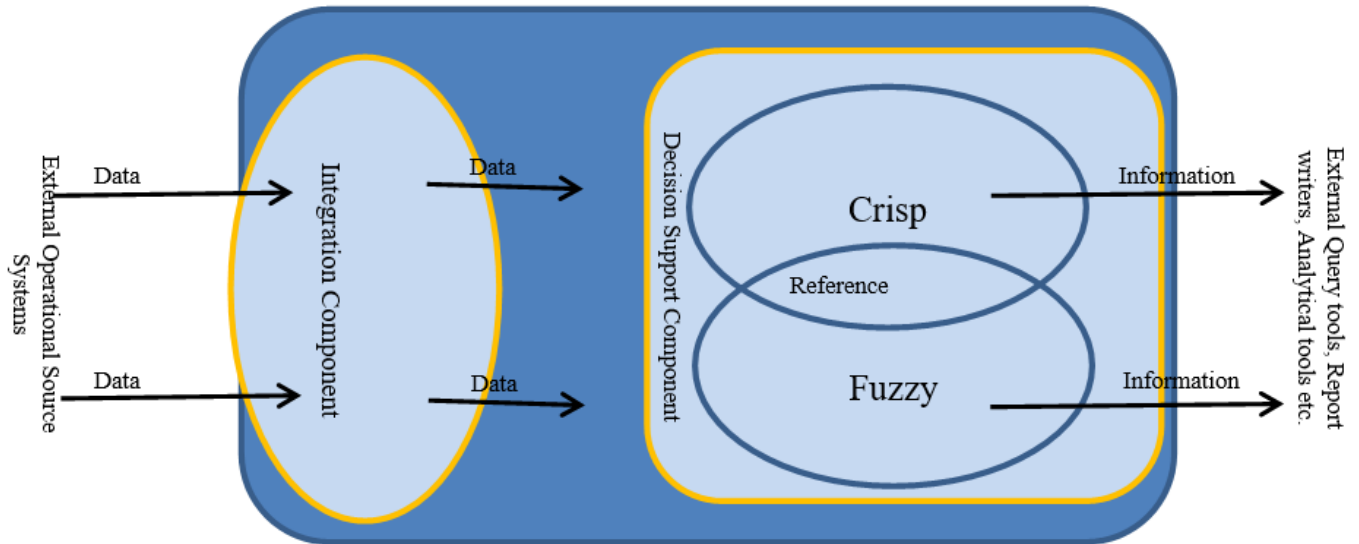


Fig. 2: Schema of the proposed FIDR (Igbape, 2014)

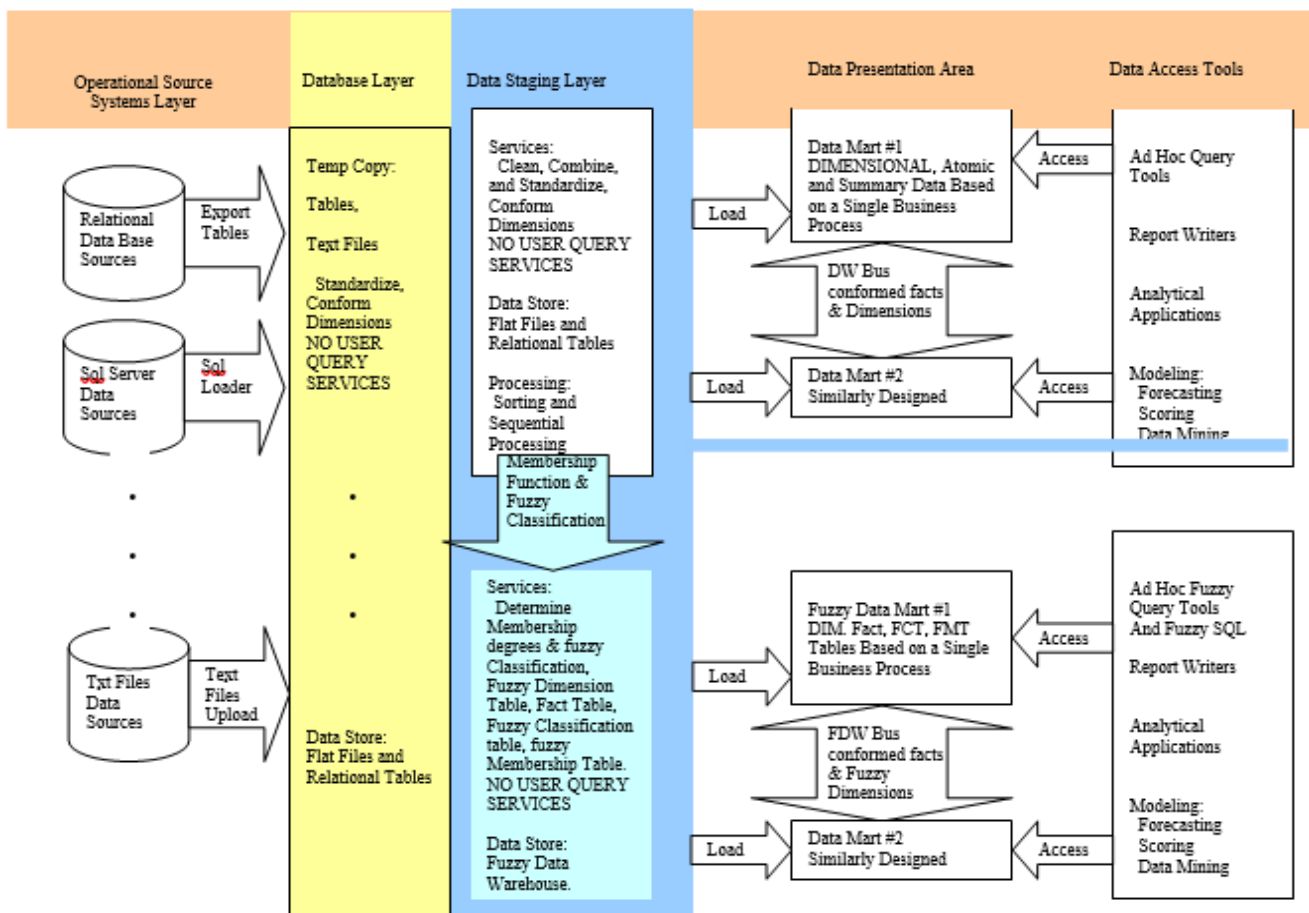


Fig. 3, Architecture for the proposed FIDR Model.

To achieve the feasibility of the above proposed extensions we adopt the definition of fuzzy data warehouse proposed by Fasel & Shahzad (2010) as follows:

A fuzzy data warehouse model is a set of tables and it is represented by FDWH.

$$FDWH = \{Dim, Fact, FCT(SA), FMT(SA)\}$$

While using definition above to achieve an extension to Kimball's model, the model proposed here has introduced a database layer to cater for integration services from different data sources.

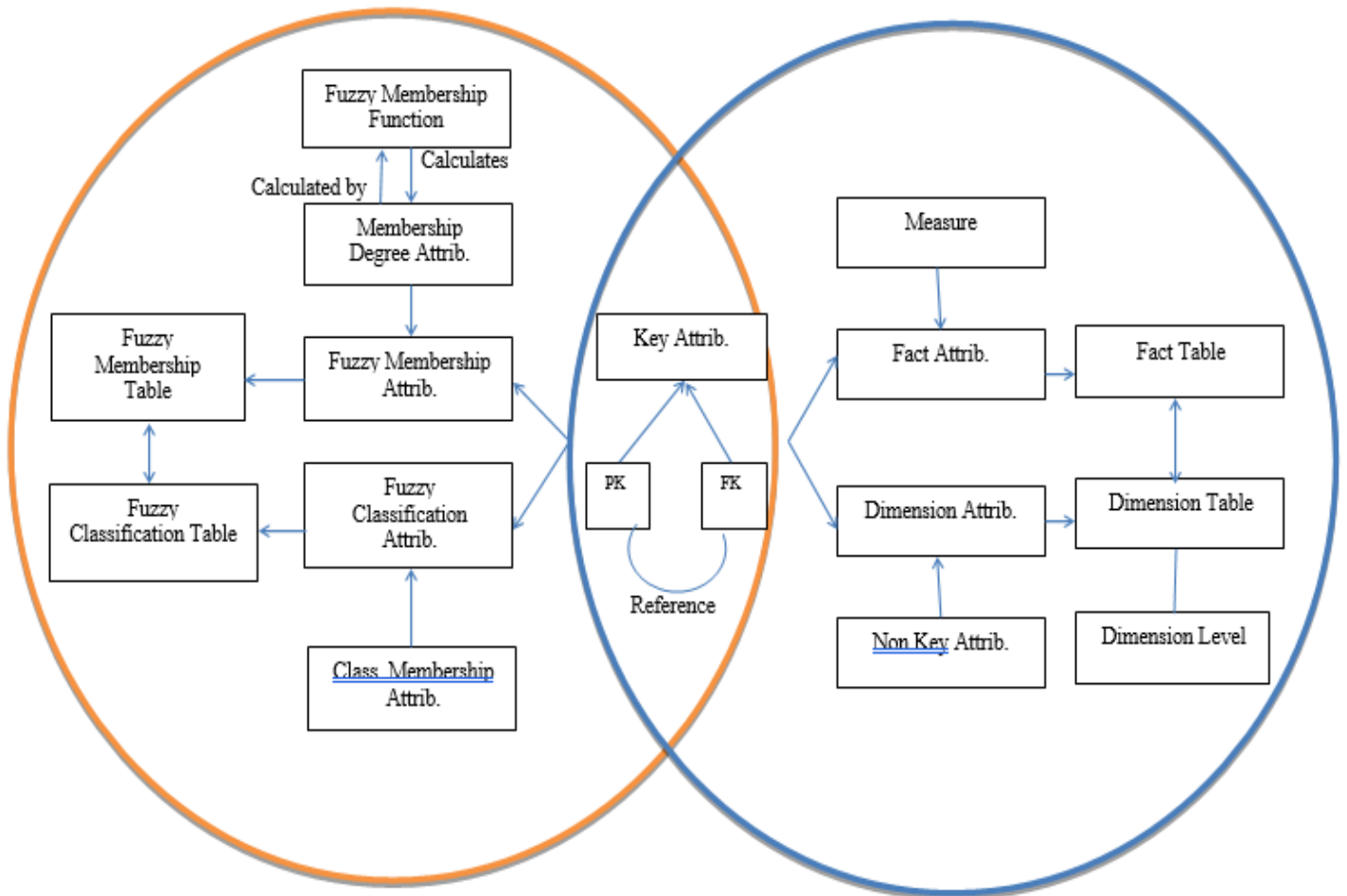
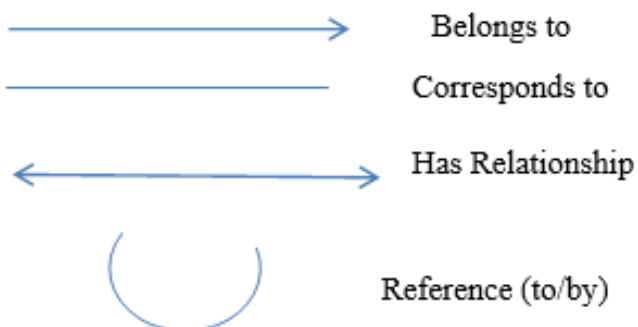


Fig. 4: Proposed Fuzzy Integrated Data Repository Meta model



FUNDAMENTAL MODELLING CONCEPTS

Let x be an attribute or atomic data having a range of possible values. The **domain of x denoted $D(x)$** is the set or range of possible values that x can have.

An **attribute x** that is required to be classified fuzzily is referred to as target attribute denoted by Tx . Instances of attribute Tx are classified over a set (S) that is represented by a linguistic variable which is a set of non-numeric terms called linguistic terms.

An attribute that has a set of linguistic terms to which Tx may belong is referred to as **Class Membership Attribute denoted by $CMx(Tx)$** . These are a set of distinct classes to which Tx may belong.

The **Membership Degree of an attribute is MDx** is an attribute that has a membership degree of Tx and is calculated by a membership function represented by $\mu(Tx) = MD$, where MD is the membership degree of Tx .

We can therefore define a fuzzy set CMx in the domain of x , $D(Tx)$ as

$$CMx = \{(Tx, \mu_{CMx}(Tx))\}$$

Where $Tx \in D(Tx)$, $\mu_{CMx}: CMx \in [0,1]$ is a membership function of CMx and $\mu_{CMx}(Tx) \in [0,1]$ is the membership degree.

The quality-assured dimensional tables in the data staging area contains measured crisp values. To achieve the requirement of realizing a fuzzy dimensional table that is mapped to the crisp dimensional table in a 1:1 relationship, two tables are used for classification as defined by Fasal and Shahzad, (2009).

Fuzzy Classification Table is a two-attribute table that consists of fuzzy classes and their unique identifiers made up of an identity attribute and a class membership attribute.

$$FCT(Tx) = \{Identifier, CMx(Tx)\};$$

Fuzzy Membership Table is a table that stores values representing the degree to which a value is related to a fuzzy class. The table has four attributes: identity attribute of the table, the identifier of the target attribute Tx , the identifier of the fuzzy classification table $FCT(Tx)$ and membership degree attribute MDx for Tx .

$$FMT(Tx) = \{Identifier, Identifier of Tx, Identifier of FCT(Tx), MDx(Tx)\}$$

The above model represented as FIDR presents a set of tables denoted by

$$FIDR = \{Dim, Facts, FCT(Sx), FMT(Sx)\}$$

Where

$$Dim = \{a \text{ set of attributes, level of attributes}\}$$

$$Facts = \{a \text{ set of facts, a set of measures}\}$$

$$Sx = \{Tx_1, Tx_2, \dots, Tx_n\}$$

Where n is the number of attributes that can be classified fuzzily.

As noted by Daniel Fasel and Khurram Shahzad, 2009, Sx is a subset of Dimension and set of facts.

Therefore;

S_x is a subset of $Dim \cup Fact$ Table (ie for all $Tx_i \in Dim \cup Fact$; i is from 1 to n)

For each attribute Tx_1, Tx_2, \dots, Tx_n

$FCT(Tx_i) = \{Identifier, CMx(Tx_i)\}$;

Where $i = 1$ to n .

$FMT(Tx_i) = \{Identifier, Identifier\ of\ FCT, Identifier\ of\ Tx, MDx(Tx_i)\}$

Where $i = 1$ to n .

DETERMINATION OF MEMBERSHIP DEGREES

Zadeh proposed a series of membership functions that could be classified into two groups: those made of straight lines or “linear” and Gaussian forms or “curves” (Zadeh, 1992)

The type of representation of the membership function depends on the base set. If the set consists of many values, or is the base set a continuum, then a parametric representation is appropriate. For that, functions are used that can be adapted by changing the parameters. Piecewise membership functions are preferred because of their simplicity and efficiency with respect to compatibility. These are mostly trapezoidal or triangular functions which are defined by four and three parameters respectively as shown below.

$$f(x; a, b, c, d) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ 1, & b \leq x \leq c \\ \frac{d-x}{d-c}, & c \leq x \leq d \\ 0, & d \leq x \end{cases}$$

This migrates into a triangular function for the case $b=c$

The quality-assured dimensional tables in the data staging area contains measured crisp values. Survey questionnaires on Likert Scale (Siti and Azmin, 2020) would be used to elicit data on attributes that show potential for imprecision and vagueness. In this study, a fuzzy classification algorithm is developed using fuzzy trapezoidal membership function and/or triangular function that will determine membership degrees of each attribute in the quality assured dimensional table and produces a fuzzy membership table that is associated with the crisp dimensional table in a 1:n or n:1 relationship.

The measured crisp input X is transformed into a fuzzy set V_x which is characterized by a membership function $\mu(Tx) = MD$ and is labeled by linguistic terms such as very good, good, average, poor, very poor and bad performance respectively. $\mu(Tx) = MD$, where MD is the membership degree of Tx . We can therefore define a fuzzy set CM_x in the domain of $x, D(Tx)$ as;

$$CM_x = \{(Tx, \mu_{CM_x}(Tx))\}$$

Where $Tx \in D(Tx)$, $\mu_{CM_x} \cdot CM_x \in [0, 1]$ is a membership function of CM_x and

$\mu_{CMx}(Tx) \in [0,1]$ is the membership degree

ARCHITECTURAL FRAMEWORK FOR IMPLEMENTATION.

In this section we provide an architectural framework for the implementation of FIDR. With the development of a frontend application, users may not be required to possess knowledge of the database structures and meta-tables. This development is not covered in this work. We therefore outline fundamental processes and an architectural framework for developers who intend to use this model to develop application systems.

While we recommend PostgreSQL 8 for Windows (Blum, 2007) this architecture will work with Microsoft SQL Server and MySQL database engine as well. It is to be noted that the database engine adopted will determine the specific syntax of the SQL commands. We have adopted the three-tier architecture framework built with PostgreSQL (Fasel, 2012).

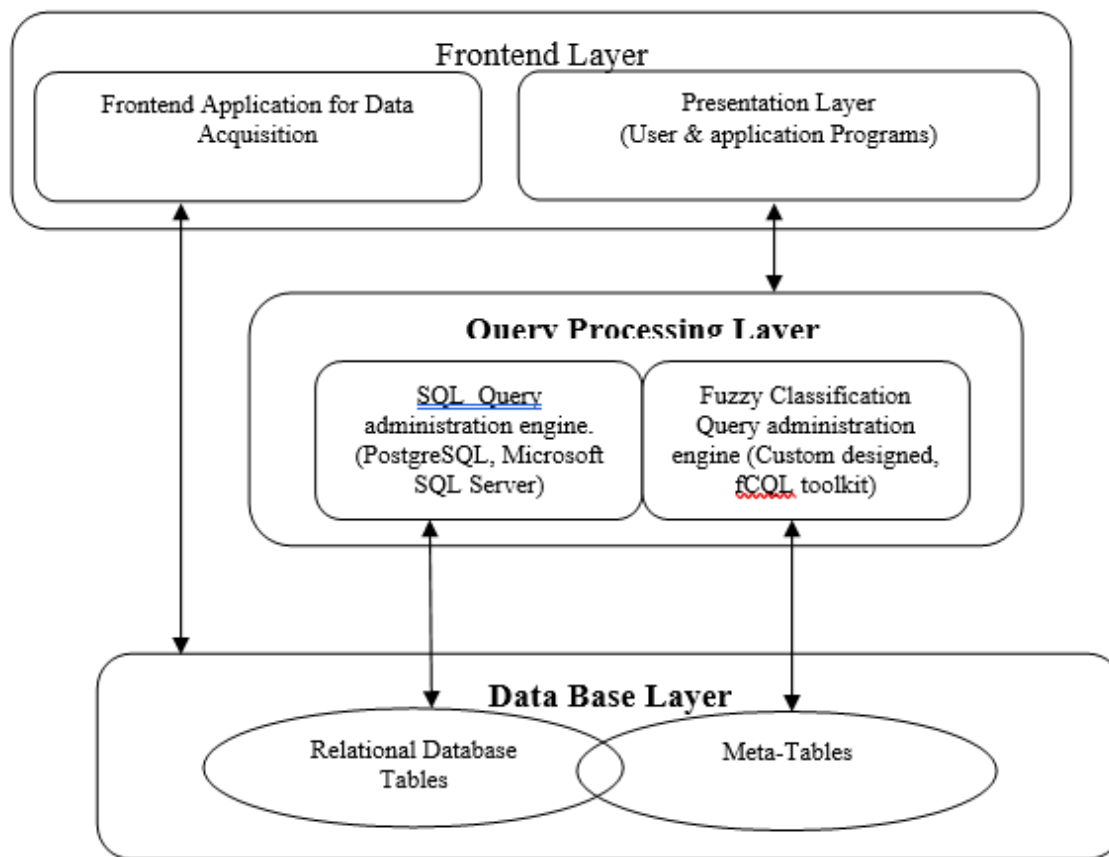


Fig. 5: Implementation Framework for FIDR

The system is achieved by extending the relational database schema through the addition of meta tables to the system catalogue. The meta tables contain definitions of linguistic variables and terms as well as description of classes and information regarding membership functions and determination of membership degrees. The business data and the meta tables are independent of one another so updates to meta tables do not affect the integrity of the business data.

Developers can create frontend applications for the query processing layer using technologies provided by the .NET development environment. These include the visual studio.NET suite and the .NET framework Software Development Kit (SDK). PostgreSQL also provides development library for .NET programmers

(Blum, 2007).

However, it is possible to integrate the fCQL toolkit (Werro, 2008) into the system as the fuzzy Query processing engine. fCQL toolkit can work with both Microsoft database engine as well as PostgreSQL database engine and it is a standalone application implemented with Java.

SUMMARY AND CONCLUSION

This study examined the key issues that have to do with unavailability of information on the practice of herbal medicine in Nigeria. It is noted that a large percentage of people in Nigeria, especially under-served and rural dwellers resort to the use of herbal medicine as their primary source of healthcare. However, there is a dearth of correctly documented information on herbal medicine practice in Nigeria. This lack of documentation of herbal medicine practice and medicinal plants constitutes a major setback in healthcare delivery, research and development and synergy with orthodox medical practice in Nigeria. This study has presented an architectural framework for the development of FIDR that can resolve the information impasse in herbal medicine and consequently facilitate access to herbal medicine information as well as preserve the traditional medicine heritage of the diverse cultures in Nigeria. This work has been able to x-ray the key components of FIDR as well as outline the procedure for integration of fuzzy concepts into the repository for effective classification of medicinal plants.

This work is the first part of a major research effort to build FIDR for medicinal plants and herbal medicine practice in Nigeria. It is hoped that the framework presented in this work will elicit the interest and curiosity of health practitioners and other stakeholders in traditional herbal medical practice in Nigeria and cross-border.

REFERENCES

1. Blum, R. (2007) PostgreSQL 8 for Windows. McGraw Hill Book Company, New York.
2. Galindo J., Medina J.M., Pons O. and Cubero J.C. (1998). "A server for Fuzzy Queries" T. Andreasen, H. Christinasen and H. L. Larsen (Eds.): FQAS '98, LNAI 1495, pp 164-174. Berlin: Springer-Verlag.
3. Fasel D, and Shahzad, K. (2010). A data warehouse model for integrating fuzzy concepts in meta-table structures. In 17th IEEE International Conference and Workshops on the Engineering of Computer Based Systems, pages 100-109.
4. Fasel Daniel (2012). Concepts & Implementation of a Fuzzy Data Warehouse. PhD Thesis submitted to the Faculty of Economics & Social Sciences, University of Fribourg, Switzerland.
5. Igbape E. M. and Enoyoze G. E. (2019) Preservation and Propagation of African Trado-Medical Heritage Using Information Technology (IT) Tools. In 1st INTERNATIONAL CONFERENCE AND EXHIBITION 2019, Samuel Adegboyega University, Ogwa, Edo State, Nigeria. Organised by College of Basic and Applied Sciences, Samuel Adegboyega University Ogwa, In Collaboration with Pax Herbal Clinic and Research Laboratories, Edo State, Nigeria. June 6-7, 2019.
6. Jin R, Zhang B, Liu XQ, Xue CM, Zhang Q, Li LZ, Zhao Q, Li K. (2012) A theoretical and experimental study on the Fuzzy evaluation model of biological performance of Chinese materia medica with an either cold or hot herbal property. Zhong Xi Yi Jie He Xue Bao. 10 (10):1106-19. Chinese. doi: 10.3736/jcim20121007. PMID: 23073194.
7. Prade, H. (1984). Lipski's approach to incomplete information databases restated and generalized in the setting of Zadeh's possibility theory. Information Systems, 9, 27-42.
8. Prade, H. & Testemale, C. (1984). Generalizing database relational algebra for the treatment of incomplete/uncertain information and vague queries. Information Sciences, 34, 115-143.
9. Siti Siryani Sofian and azmin Sham Rambely (2020). Measuring Perception of Students towards Game Recreation activity using Fuzzy Cojoint Analysis. Indonesian Journal of Electrical Engineering

and Computer Science Vol.20, No.1.Oct. 2020, pp 395-404.

10. Werro, N. (2008). Fuzzy Classification of Online Customers. PhD Thesis submitted to Faculty of Economics and Social Sciences, University of Fibourg, Switzerland.
11. Zadeh L. A. (1994) Fuzzy Logic, Neural Networks and Soft Computing: Communication of the ACM, March 1994, Vol. 37, No. 3 pp 77-84.
12. Zadeh L. A. (1965) Fuzzy sets. Information and Control, 8, 338-353.
13. Zadeh L. A. (1978) Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems, 1, 3-28.
14. Zadeh L. A. (1992) Knowledge representation in fuzzy logic: An introduction to fuzzy logic applications in intelligent systems. Kluwer Academic Publisher.
15. Zimmermann, H. J. (1991). Fuzzy set theory and its applications (2nd). Kluwer Academic Publishers.