EXTENDING THE REACH OF KNOWLEDGE ENGINEERING PRACTICE BEYOND THE FRONTIERS OF FORMAL SECTORS USING A FEATURE AND FUNCTION BASED CLASSIFICATION

ODULAJA, Godwin Oluseyi
Computer Science Department, School of
Computing and Engineering Sciences
Babcock University,
Ilishan Remo, Ogun State Nigeria

AWODELE, Oludele
Computer Science Department, School of
Computing and Engineering Sciences
Babcock University,
Ilishan Remo, Ogun State Nigeria

IDOWU Sunday A.
Computer Science Department,
School of Computing and Engineering Sciences
Babcock University,
Ilishan Remo, Ogun State Nigeria

NWAOCHA, Vivian O.
Computer Science Department,
National Open University of Nigeria
Ilishan Remo, Ogun State Nigeria

Abstract: In recent years, there had been many frameworks on improving Knowledge Engineering (KE). These frameworks had emerged for reasons ranging from shortening KE product development time, performance improvement, to addressing knowledge acquisition challenges. Despite these, impact of KE products (Expert Systems (ES) and Knowledge Based Systems (KBS)) has not been felt in non-formal sectors where there is no formalized way of keeping and exchanging their specialized knowledge. Here, heuristics and implicit tacit knowledge are often used to solve significant problems. When custodians of such specialized knowledge die, their untapped problem solving skills perish with them. Such is common in Africa, especially among native farmers, hunters and healers. Extending the reach of KE practice to these sectors can effectively help overcome this challenge. This however requires a better understanding of existing KE frameworks in a bid to isolate gaps responsible for this oversight and factor solution to this problem into subsequent design of a KE framework. To this end, this paper provides a concise feature and function based classification and review of eight prominent KE frameworks and models (CommonKADS, MIKE, MOKA, PROTÉGÉ II, SPEDE, RLM, CRLM, and PGM/CPGM) in common use in recent years and consequently recommends development of a KE framework that addresses the isolated problem.

Keywords: Expert Systems, Non-formal Sectors, Knowledge Engineering, Specialized Knowledge, Domain Knowledge Custodian

I. INTRODUCTION

Nigeria and many other nations around the world are rife with native and indigenous life-enriching skills, cultures and solutions to many ailments and social ills. Custodians of these special forms of knowledge include traditional healers, game hunters, warriors, farmers and certain families often found in rural and remote areas in whose lineage the knowledge is domiciled. The problem however is that largely, individuals, families and organizations in possession of this specialized knowledge and skills often lack a formal way of documenting, storing, preserving, and exchanging this knowledge other than the oral and practical demonstrations passed down from one generation to another. A few others at best strife to preserve these knowledge and procedures using manual paper and pen approach. However, factors such as economic instability, social unrest, modern civilization drive, natural disasters, and search for western education in the more civilized cities and economies, and the likes often lead to internally displaced people (IDP), death or migration to better economies. As these factors separate up-coming generations from their progenitors who are the custodians of the specialized knowledge, this inevitably results in loss of

some or all of this knowledge with appalling consequences especially when such knowledge custodians die.

However, Knowledge Engineering (KE) which is a subfield of Artificial Intelligence that focuses on design and development of Expert Systems (ES) and Knowledge Based Systems (KBS) as a technological intervention can be used to abate this trend. KE products are software applications specially designed and developed to carry out the role of a human expert, as a substitute or in the least, to support the human expert in executing his cognitive prowess [1]. Application sectors include health, transportation (autonomous cars) [2], security, commerce and industry, agricultural, social media, aerospace and the Internet.

These products have benefitted man in the formal sectors in many critical and life crucial specialist fields for decades, and their reach can be extended to non-formal sectors for good [3]. Poor adoption of technology for instance had been cited as one major cause of food (cocoa) production problems in Africa [4]. Although methodologies for building KE products are many and proliferated, impact of KE has not been duly felt in several non-formal organizations and this calls for urgent attention [5]. As this

trend perpetuates, important domain and specialized knowledge continue to escape being captured and harnessed for the benefit of mankind escape with. This is especially true when the custodians of such knowledge die. In a bid to appropriately remedy this problem, this study is poised at first establishing factors responsible for this problem and then recommend appropriate solution.

II.   REVIEW OF LITERATURE

In literature, many problems had been cited as challenges facing KE. These include poorly managed knowledge base, products with integrity issues [6], inconsistent query processing techniques and lack of consistent data organization and exchange processes [7]; products that failed to meet changing user requirements specifications as new technologies (such as mobile and ubiquitous computing) emerge. Smith [8] emphasized the need to address knowledge elicitation problems in order to avoid deadlock. Paucity of and lack of cooperation from domain knowledge experts and providers [9]; [10]. These people admittedly are few and on high demand. This led to challenges with populating and effectively managing the knowledge base of KE products.

Ubiquitous Computing - In recent years, demands for KE products that satisfactorily meet users' needs at their comfort zone and time are on increase. This can be below expectation given the power and ubiquity of internet technology and level of proliferation of models, frameworks and products of KE circulating around the world Failure to factor emerging technologies into design and implementation of KE products (intelligent systems) is directly responsible for this.

Sector Oversight - Individual and private sectors without a formalized way of keeping and exchanging knowledge suffer. Many such ones use heuristics and tacit knowledge to solve important problems, sometimes without following any sequential procedure [4]. When such solution providers as these die, their untapped problem solving skill perishes with them and remains inaccessible to the world. Such is common in Africa, especially among native farmers, hunters and traditional healers.

In-appropriate Tools and Technology - Prolonged attachment to outdated development tools and techniques [7] has resulted in interface usability challenges [4] and other design issues which in turn had resulted in low intake of technological innovations, imprecise and poor knowledge transfer processes, inadequate extension and advisory services and other negative, diverse and sector-dependent implications. Hence the need for improved method of developing KE applications that captures the needs of formal and non-formal sectors alike [6].

III.   RELATED WORKS

Atanasov and Krupka [12] in an ES design called Modern Expert System Architecture, added new functional features such as interactive module, coordination module, rule editor and knowledge editor intended to improve ES design and performance further. They also tried to separate

knowledge acquisition (gaining knowledge) from the core ES shell.

As a reference point, in the industry cluster, the unresolved issues of CommonKADS had, in some cases in practice called for replacement of these models   (with cluster model, collaboration models and combined *scenario*, *architecture* and *specification* models, respectively, (compare fig. 1 and 2) because of lack of concrete definition in its design model. This lack could cause the knowledge engineer to misinform the system developer [5].
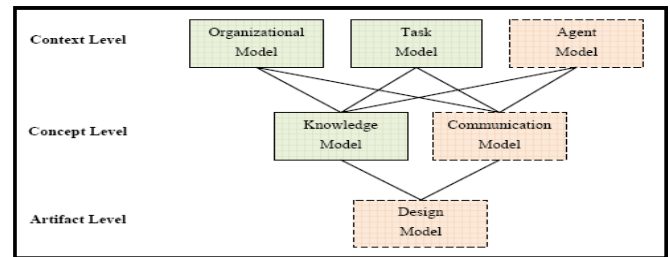

Figure 1: CommonKADS Model Suite
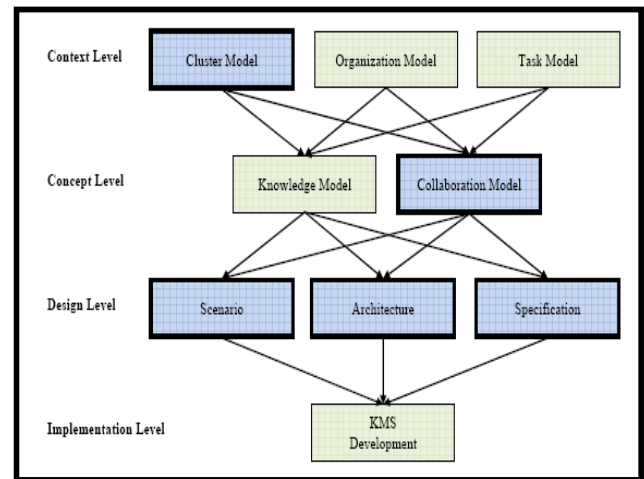*(Source: Sureephong et al, 2016)*

Figure 2: Proposed Methodology for Developing KMS for the Industry Cluster
(Source: Sureephong et al, 2016)

Shervan, Hadi and Shahram in [13], developed a decision support expert system to automate employment of university course lecturers. It was a rule based expert system that uses a decision tree data mining and majority voting algorithm to make selection among contending applicants based on qualification, work experience and area of specialization. It was evaluated to have 78.88% accuracy on average. It however is bound to depend on too many rules in the knowledge base which is inevitable as the application domain expands, very complex algorithm that requires longer turn-around time will be required to manage the inference engine. Thus it will crash in a short time as several university departments are added or in the least become ineffective.

Atanasova and Krupka [12 in their Modern Expert System Architecture reflected the presence of new elements such as interactive module, coordination module, rule editor and knowledge editor in design of modern ES.

Saket, Akankasha and Vikas in [14] explored the structure and role of expert system in agriculture and gave an overview of feasibilities of designing, developing and implementation of expert systems for agriculture. Their objective is to motivate scientists and extension workers to peer into possibilities of developing applications for the benefit of entire global agricultural community. The work however lacks specific implementation framework that factors emerging technologies into ES design.

Diana-Aderina in [**15**] specified that expert system components have gone beyond consisting of just the traditional three components: inference engine, user interface and knowledge base and there is the need to acknowledge this design of modern ES. This opinion supports the work of Turban, Sharda and Delen, (2011) who added the workspace, the application module and the knowledge refinement module to the structure of expert system (Turban, Sharda and Delen, 2011) as distinct and prominent features of the knowledge acquisition subsystem.

Wiliyanto, [16] in a study conducted by Dian Atnantomi Wiliyanto to determine the effectiveness of web based expert systems application in identifying and intervening the lives of children with special needs, he found that use of expert system is effective in identifying and intervening to help children with special needs in inclusive schools. She recorded 52% result accuracy.

Santosa, Romla and Herawati, in [17] used Fuzzy Mamdani, Santosa, Romla and Herawati to develop an expert system to detect and care for cataract eye disease based on the symptoms identified. From a sample of 50 cases, the software had 78% prediction accuracy compared with that of human experts.

Shodhganga [6] compared several ES development models applicable in Agric and came up with a web based model. He pointed out that ES is an expertise driven research area where researchers have many times used multi methodologies approach to develop ES.

Significantly, these studies neither address the need to extend KE to non-formal sectors, nor establish why the oversight had been sustained. Hence, the need for a feature and function based classification and analysis of prominent KE methodologies in recent years.

IV. A FEATURES AND FUNCTION BASED CLASSIFICATION OF KNOWLEDGE ENGINEERING FRAMEWORKS AND PROBLEM SOLVING METHODS (PMS)

Table 1 gives a feature and function based characterization and classification of some prominent selected knowledge engineering methods developed in the past few decades. This was carried out to determine what existing gaps need to be filled in order to address this peculiar application oversight challenge. Factors considered are: the *basic underlying assumption(s), principles or* *strength of the methodology; its application domain or scope; object orientation capability; implementation complexity; integration of knowledge elicitation and acquisition facilities and; identified gaps or weak points in literature* if there is any.

A. CommonKADS / KADS – II

This is the European de facto standard for knowledge analysis and development of knowledge intensive systems. It is strongly model based, equipped with powerful modeling reusable templates for knowledge management, analysis and application development. Its six models – *organizational model, task model, agent model, knowledge (or expertise) model, communication model and design model* capture specifics of the KBS to be built as well as its proposed organizational operating environments. While the *expertise and design models* focus on describing functional and non-functional aspects of the KBS under construction, the other four models focus on modeling the organization where the KBS will be used and the tasks carried out within the organization.

Although, this model has been acknowledged as the industry standard, it can however be argued that CommonKADS was developed for *well defined formalized "hard" domains w*hereas the industry is a "semi-soft" domain. This implies that necessary modifications to match organizational context and environment are required. In CommonKADS for instance, the *Agent* model specifies what task is carried by particular agent, what other agents are involved in executing a task, qualification of each agent, list of tasks for each agent and knowledge items possessed by each agent. This is not well defined in industry context though because this concept is based on the assumption that communication in the industry context is intra and horizontal such that tasks assignments to agents are internal and stereotyped. Reality however dictates otherwise. There is industrial interdependence of knowledge and resources and knowledge communication must be both intra and inter, vertical and horizontal (a shoe and leather bag factory for example depends on and must communicate with hide and skin processing industry) to achieve production objectives. This weakness of the agent model by extension directly and adversely affects the communication model whose role it is to model knowledge exchange between agents for the industry. By the same token, the design model is adversely affected. Therefore, this model require modification so as to make expected output clear and unambiguous (Sureephong, 2016; Sandberg, 96).

B. MIKE

Model-based and Incremental Knowledge Engineering, MIKE is a model based problem solving method (PSM) that strongly relies on *KARL* (a formal operational knowledge specification language) for post analysis knowledge acquisition and test-running and on *DesignKARL* language for design implementation. Its five cyclic phases are: *Knowledge Elicitation, Interpretation, Formalisation and Operationalisation, Design and Implementation.* MIKE improved on CommonKADS by supporting prototyping and incremental reversible system development. Another strength of MIKE is associated with the fact that KARL model in MIKE is executable. This makes validation of the

expertise model and involvement of the expert in the validation process via prototyping possible.

MIKE uses structured interviews to acquire expert knowledge and stores these in natural language in the informal so called **knowledge protocols.** Structuring and formalization phases transform this to the *KARL model* with all functional requirements. Non-functional requirements are added at the iterative design phase to generate the *design model.* The language DesignKARL enhances KARL features with ability to describe data types and algorithms as well as possession of features needed to structure KARL model further. It also made description of design process and associated decision making operations possible. This culminates in the implementation phase which is done on target hardware and operating system.

However, MIKE as shown is limited in application scope to modeling KBS under development as it does not scale well to capture all that it takes to embed KBS into big business environment. It requires extended models that can capture different views of the business environment.

C. Other Models

PROTÉGÉ II, MOKA, PGM/CPGM, SPEDE, Role Limiting Methods (RLM) and Constrained Role Limiting Methods (CRLM) are some other PSMs developed for KBS development over the past few decades other than CommonKADS and MIKE. Table 1 gives a feature and function based classification of eight prominent knowledge engineering methodologies developed in the past few decades.

Table 1: Features and Function Based Classification of Some Prominent Knowledge Engineering Approaches

Sr./No.	KE FRAMEWORK MODEL OR METHODOLOGY	BASIC UNDERLYING ASSUMPTIONS / PRINCIPLES / STRENGTH	APPLICATION DOMAIN	OBJECT ORIENTED (OO) COMPATIBILITY	IMPLEMENTATION COMPLEXITY	INTEGRATION OF KNOWLEDGE ELICITATION & ACQUISITION FACILITIES	GAPS /WEAK POINTS
1	**CommonKADS / KADS – II** – the European de facto standard for knowledge analysis and development of knowledge intensive systems	Knowledge modeling, management, storage, documentation and system development	Generic for business organizations, corporate systems and industries	OO and UML Compatible	Pre-implementation system analysis and modeling are intensive but implementation is relatively fast and easy	Not well defined. Emphasis is on modeling, structuring and documentation of knowledge and procedures	Knowledge elicitation facility is not well defined Three models (Agent, Communication and Design) out of its six are not well defined for industry cluster
2	**PGM / CPGM – P**robabilistic **G**raphical **M**odel / Canonical probabilistic graphical Model	For representing complex domains with the use of probability distributions. Based on either Bayesian Directed Acyclic Graph (DAG) network or Markov undirected graphical model / Networks	Not generic, found applicable to machine learning (ML), Natural language Processing (NLP), Computational Biology (CB), Computer Vision (CV).	OO and UML Compatible	There is generally shortage of knowledge of how Canonical PGM works, what they express and when they can be used. There is not enough literature to guide a novice knowledge engineer.	Emphasis is on knowledge representation, not on acquisition or capture which is not well defined in literature. Availability of knowledge or data needed is taken for granted	Ambiguity problem has made classification of isolated data points independent of each other a bane of Machine Learning, Natural Language Processing, and Computer Vision. CPGM that could help is not well articulated in literature.
3	**MOKA** – **M**ethodology and tools **O**riented to **K**nowledge-Based Engineering **A**pplications	Knowledge Reuse, reduced turn-around time and platform independent implementation.	Specifically developed for aeronautical and Automotive industries.	MOKA is OO Compatible. Fosters knowledge reuse	Compatible with any KBE Platform Adoption of formal and informal models. ICARE forms for informal model and platform independent pseudocode for formal model	Knowledge elicitation is not well defined. But it has ICARE forms as templates for knowledge acquisition.	Not generic, but specific to engineering task domain.
4	**MIKE** – Model-based and Incremental Knowledge Engineering	It is model based, and strongly relies on KARL (a formal operational knowledge specification language) for post	Modeling KBS under development	MIKE is OO Compatible.	Since KARL used for formal operational knowledge specification is an executable language, this allows the expertise model to be built by	Uses structured interviews for informal knowledge elicitation, and knowledge protocols for	Limited in scope. Requires extending it with more models to be able to capture embedding of

#							
		analysis phase and DesignKARL language for design implementation phase. Tried to leverage on CommonKADS by supporting prototyping and incremental reversible system development.			explorative prototyping. Ditto to Design Model which is iteratively refined and run using DesignKARL until satisfactory.	storing knowledge expressed in natural language	KBS into business environment to capture different enterprise views. Since KARL (an interpreted language) cannot be used to specify real time problems, this limits scope of MIKE too.
5	**PROTÉGÉ-II** – A suite of tools (MAITR & DASH) that automate building knowledge acquisition tool given abstract models (domain ontologies) of the proposed system	Rapid Development of KBS using special tools. MAITRE tools helps builders create and build domain specific ontologies of target application. DASH tools automatically build a knowledge acquisition shell/tool into which domain specific knowledge is entered by the knowledge engineer.	Building Intelligent systems, Knowledge Based Systems, Knowledge Engineering applications	PROTÉGÉ II is OO Compatible and promotes reuse of components and procedures	Given that the domain ontologies (content knowledge) are available, the suite of tools (MAITRE, DASH etc) automate building required application	Elicitation of domain knowledge is not captured in its design.	The domain dependent knowledge (ontologies) has to be sourced and keyed into the knowledge acquisition tool built by PROTÉGÉ II by the knowledge engineer. Still undergoing design improvements & requires large projects to fulfill its goal.
6	**SPEDE** -Developed by Rolce-Royce with consultation support from Epistemics	Developed for use in knowledge management to help businesses capture validate and communicate vital information.	Medium to Large Commercial Business Domain	SPEDE is OO compatible.	Implementation made easy with the 10 tools embedded in PCPACK5. The series of review meeting of knowledge engineers and management (tagged gates) aids in a spiral implementation.	Knowledge elicitation from domain experts remains a major challenge. There are series of knowledge modeling and classification templates however.	Not generic. Difficult to implement in non formal / non-business knowledge domains where knowledge elicitation and validation from the domain experts must precede electronic data capture. Process not being autonomous, knowledge engineer(s) spend much time to key domain knowledge.
7	**Role Limiting Methods (RLM) e.g. SALT (a Propose-and-Revise PSM)** – a shell approach Problem Solving Method (PSM) that fosters reusable problem solving methods.	Uses role limiting shells and predefined structured graphical user interfaces to guide and ease knowledge engineers' task of populating knowledge base during knowledge acquisition phase.	Applicable to specific types of problem for which the shell is appropriate.	RLM is OO compatible	For efficiency, there is need for mapping tasks types to corresponding RLM – a difficult-to-meet requirement. Its fixed structure (shell) does not sit well with problems requiring application of several problem solving methods	Knowledge elicitation is not well defined, has inflexible shell-based knowledge entry graphical user interfaces given that the knowledge is available.	Rigidity, brittleness inflexibility. It is only applicable to specific types of problem for which the shell is appropriate. Its predefined knowledge acquisition structures constraint experts during knowledge acquisition

8	CRLM – Configurable RLM. An approach to add flexibility to RLM by allowing for several PSMs (such as Heuristic Classification and Set-Covering Classification) to run within the same shell via reconfiguration of the shell.	Decomposition of Shell's complex PSM into several subtasks each of which several solution methods are allowed and applied	Applicable to wider range of problems for which the shell is appropriate with overheads.	CRLM is OO compatible	Configuring and re-configuring the CRLM burdens the system developer each time.	Knowledge elicitation is not well defined, it has inflexible shell-based knowledge entry graphical user interfaces given that the knowledge is available.	Every CRLM has a fixed scheme of knowledge types and predefined communication paths. Configuring the CRLM burdens the system developer each time.

V. RESULTS AND DISCUSSIONS

Findings show that despite the presence of knowledge modeling, management, storage, documentation and refinement features present in almost all these analyzed KE approaches, 87.5% of them did not have well defined features for knowledge elicitation. Building knowledge modeling, management, storage, documentation and refinement templates should be preceded with knowledge elicitation since only available knowledge can be managed, modeled or refined. Thus Knowledge acquisition problem, which is not just about modeling, management, storage, documentation and refinement of knowledge but importantly includes knowledge elicitation which had largely been overlooked in literature, had predominantly been responsible for poor impact and performance of KE products in informal sectors. It is imperative to devise an effective means of eliciting knowledge from all in the knowledge domain, be it formal sectors (whose domain knowledge might be already well structured, in electronic and easily convertible form) or non formal sectors (whose knowledge exists in unstructured, non electronic, hand-written, and sometimes oral form). This might require tact, patience, motivation and strong persuasiveness (convincing skill) on the part of the Knowledge Engineer in order to convince them and gain their cooperation. It is therefore recommended that KE frameworks that factor knowledge elicitation, (not just knowledge acquisition models and refinements) into design of KE product and development be developed and adopted for KE product development and practice.

VI. CONCLUSION

This study identified knowledge acquisition (knowledge elicitation to be precise) bottlenecks as the major factor responsible for lack of significant impact of KE practice in informal sectors around the world but especially in Africa. It also discovered that while there had been numerous knowledge acquisition features in almost all existing KE frameworks and methodologies, emphasis had been on knowledge refinements and modeling, and structuring knowledge capture templates while knowledge elicitation receives little attention. Other problems emanating from this include poor KE product design and performance validation. Since knowledge elicitation must precede knowledge capture, acquisition onto electronic templates and refinements, this should receive adequate and immediate attention. Extending the reach of KE to non formal sectors therefore require factoring knowledge elicitation into the design of KE products. This study therefore recommends design and development of a framework in KE that factors knowledge elicitation from all necessary sectors, formal or non formal into its design. This will enable Knowledge Engineers to salvage and prevent important but inherently tacit and heuristic knowledge domiciled with individuals such as farmers, hunters and traditional healers found in non formal sectors in African and some other remote places in the world from perishing with custodians of this specialized knowledge when they die. It will also help solve the associated problems of KE product design and performance validation.

VII. REFERENCES

[1] M. Rouse, and M. Haughn, (2017) Definition of Knowledge Engineering. Retrieved from https://searchenterpriseai.techtarget.com/definition/knowledge-engineering

[2] Gibbs, S.,l (2016). Autonomous Cars – Who is building them and How. Retrieved from https://www.theguardian.com/technology/2016/may/26/self-driving-cars-whos-building-them-and-how-do-they-work

[3] S. Rudi, V.R. Benjamins, and F. Dieter (2017). Knowledge Engineering: Principles and Methods. Retrieved from https://www.researchgate.net/profile

[4] J. Anga, (2014) Production of cocoa beans, Quarterly Bulletin of Cocoa Statistics, London, 2006, XXXI, XXXIX (2013), XL (2014).

[5] P. Sureephong, N. Chakpitak, Y. Ouzroute, G. Neubert and A. Bouras (2016) Knowledge Engineering Techniques for Cluster Development. Retrieved from https://www.researchgate.net publication › 301872111_Knowledge_Engin...

[6] M. Shodhganga (2018). Agricultural Expert System. Retrieved from http://shodhganga.inflibnet.ac.in/bitstream/10603/71534/10/10_chapter%203.pdf

[7] M. Kantureeva, A. Zakirova, T. Mannapova, M. Mussaif, and K. Nigmetov (2014). The Methodology of Expert Systems. 14(2).

[8] S. Smith, Expert System Architecture Block Diagram (2016) Retrieved from

https://www.google.com/search?q=expert+system+architecture+block+diagram

[9] S. Osindero, Y.W. Teh, (2006). A fast Learning Algorithm for Fast Learning Nets (PDF). Neural Computations 18(7), 1527–1554.

[10] S.I. Ele, E.E. Umoh, and W.A. Adesola, (2014). An Overview of the Development, Principles, Stages and Building Blocks of Expert System. West African Journal of Industrial & Academic Research 2(1)

[11] N. Sriram and H. Philip (2018). Expert System for Decision Support in Agriculture Retrieved from http://agritech.tnau.ac.in/pdf/14.pdf.

[12]I. Atanasova, & J. Krupka (2013). Architecture and Design of Expert System for Quality of Life Evaluation. Informatica Economică , 17(3), 28-35.

[13]F. Shervan, T. Hadi and J. Shahram (2013) Design and Development of an Expert System to Help Head of University Departments in International Journal of Science and Modern Engineering (IJISME) 1(2), 45 – 48.

[14]M.A. Saket, and D. Vikas (2014). Expert Systems In Agriculture: An Overview; International Journal of Science Technology & Engineering. 1(5).

[15]M. Diana-Aderina, & A. Mihai-Constantin. (2015) Architectural Model of Expert Systems. International Symposium Engineering and Competitiveness (EMC 2015). Retrieved from https://www.researchgate.net/publication/281268197_ARCHI TECTURAL_MODEL_OF_EXPERT_SYSTEMS

[16]D.A. Wiliyanto (2017). The use of Web based expert system application for identification and intervention of children with special needs in inclusive schools. journal.uad.ac.id › Home › 11(4)

[17]Santosa, Romla and Herawati, (2018) Expert System Diagnosis of Cataract Eyes Using Fuzzy Mamdani Method. Journal of Physics: Conference Series, 953(1).