Performance Evaluation of Fuzzy based Ontology System for Identification of Malaria Disease

Nitesh Vyas
Ph.D. Scholar, Pacific Academy of Higher Education and Research, Udaipur, Rajasthan, India

Parashu Ram Pal, Ph.D.
Director, Lakshmi Narain College of Technology-MCA, Bhopal, M.P., India

ABSTRACT

The disease identification from the symptoms and other detail is one of the important research area for researcher. The various ontology systems which relate the symptom and disease is evaluated by the researchers. Many time this system fail due to confusing symptom which make the fuzzy problem for classification. In this paper fuzzy based ontology system is presented. The fuzziness of the symptoms is solving by generating and optimizing the rule base. The Malaria disease has been used to evaluate the performance of classifier. The accuracy of classification on symptoms database is the key parameter used for performance evaluation of method.

Keywords

Fuzzy. Ontology, XML, Malaria, Symptoms.

1. INTRODUCTION

The term 'ontology' (or ontologia) was coined in 1613, independently, by two philosophers, Rudolf Göckel (Goclenius), in his Lexicon philosophicum and Jacob Lorhard (Lorhardus), in his Theatrum philosophicum. Its first occurrence in English as recorded by the OED appears in Bailey's dictionary of 1721, which defines ontology as 'an Account of being in the Abstract'[4]. In 1995, Russel and Norving defined ontology as "Ontology is a formal description of the concepts and relations which can exist in a community of agents."[7]. In 1996 Swartout defined ontology as "Ontology is hiearchically structured set of terms to describe a domain that can be used as a skeletal foundation for a knowledge base."[8]. In 2000, Fensel defined ontology as "Ontology is common, shared and formal description of important concepts."[1] The ontology is a formal conceptualization of a real world, and it can share a common understanding of this real world [11]. In 2001, Noy and McGuinness defined ontology as "An ontology is formal explicit representation of concepts in a domain, properties of each concept describes characteristics and attributes of the concept known as slots and constrains of these slots." [9] In 2002, Fonseca defined ontology as "An ontology is a theory which uses a specific vocabulary to describe entities, classes, properties and related functions with certain point of view."[10]

A conceptualization can be defined as an intentional semantic structure that encodes implicit knowledge constraining the structure of a piece of a domain. Ontology is a (partial) specification of this structure, i.e., it is usually a logical theory that expresses the conceptualization explicitly in some language. Conceptualization is language independent, while ontology is language dependent. Formally ontology is the statement of a logical theory.

Ontology defines a common vocabulary for researchers who need to share information in a domain. It includes machine-interpretable definitions of basic concepts in the domain and relations among them.Ontology learning can be made from unstructured sources (NLP techniques), semi-structured sources (XML, HTML) and structured data (databases and dictionaries).

Advantages of the ontology technique are as follows :

To share common understanding of the structure of information among people or software agents

To enable reuse of domain knowledge

To make domain assumptions explicit

To separate domain knowledge from the operational knowledge

To analyze domain knowledge

The medical ontology is an important field of ontology. There is a wealth of data available within the healthcare systems but they lack effective analysis tools to discover hidden relationships and trends in data. This is obviously a difficult task and requires a profound analysis of the structure and the concepts of medical terminologies. But it can be achieved by constructing medical domain ontologies for representing medical terminology systems. Ontologies enable high scalability in searching, extracting, maintaining and generating information. Medical ontologies are developed to solve problems such as the demand for reusing and sharing patient data or the transmission of these data. With the support of the ontology, both user and system can communicate with each other by the shared and common understanding of a domain [12][13][14]. The emerging technologies in IT sector have opened exceptional opportunities in delivering the health care services well on time effectively. This has resulted in demand for intelligent and knowledge based systems which are enabled with advanced medical practices globally [15]. Electronic Health Record provides easier clinical data analysis. By which doctor can spent more time on patient care. With the help of EHR Physicians, hospitals, specialists and providers in the circle of care can share information quickly and easily.[3]

At the practical level, ontologies can help us verify the soundness of statements in messages based on our conceptualization of the world. Information models in healthcare often take the form of an XML schema, a Schematron schema, or a relational database schema. One distinguishing characteristic of ontologies is that they are based on an Open World Assumption (OWA) which is based on the AAA slogan or Anyone can say Anything about Any topic. Statements that are not included in an ontology are considered unknown as opposed to false. In contrast, information models of data structures such as XML messages and relational databases are based on a Closed World Assumption (CWA) which holds that any statement that is not known to the message or database to be true is false (this is also referred to as "negation as failure" or NF). [18]

2. MEDICAL ONTOLOGY FOR HEALTH CARE SYSTEM

It is generally acknowledged by the scientific community that ontologies may give a significant contribution to the design and implementation of better and more interoperable information systems, also in medicine. The use of ontologies in medicine is mainly focussed on the representation and re-organization of medical terminologies. Physicians developed their own specialized languages and lexicons to help them store and communicate general medical knowledge and patient-related information efficiently. Such terminologies, optimized for human processing, are characterized by a significant amount of implicit knowledge.

Medical information systems, on the other hand, need to be able to communicate complex and detailed medical concepts (possibly expressed in different languages) unambiguously. This is obviously a difficult task and requires a profound analysis of the structure and the concepts of medical terminologies. But it can be achieved by constructing medical domain ontologies for representing medical terminology systems.

3. FUZZY LOGIC AND ONTOLOGY ENGINEERING

Expert system for taking decisions against variety of diseases applying ontology engineering techniques may give a new dimension to the health care. System of such kind may befit human kind by working with each through online. Fuzzy logic has been proved to be the remarkable tool for building intelligent decision making systems based on the expert's knowledge and observations.

In medical field, the doctors and medical practitioners exhibit a great difference in decision making because of their approaches to deal with uncertainties and vagueness in the knowledge and information about medical data of patients. The diagnostic decisions also depend upon experience, expertise and perception of the practitioner. As the complexity of system increases, it is not easy to follow a particular path of diagnosis without any mistake.

Fuzzy logic gives powerful reasoning methods that can handle uncertainties and vagueness in medical field. The Fuzzy Expert Systems (FES) define imprecise knowledge and offers linguistic concept with excellent approximation to medical texts. Fuzzy logic is a method to render precise what is imprecise in the world of medicine. FES plays an important role in medicine for symptomatic diagnostic remedies. The technocrats identified potential and possible areas for implementation of FES for medical diagnosis. Also, the efforts have been made by various researchers to establish a roadmap to forecast the future developments of expert systems in medical diagnosis. The literature published upon FES in medical diagnosis encompasses a wide spectrum including reviews, applications, innovations, conceptual studies and development of diagnostic tools. The communications between medical scientists and computer engineers have lead to an interdisciplinary advance in the development of intelligent supporting tools and systems.

The use of fuzzy logic for medical diagnosis has the history of almost three to four decades. Computer scientist, technologists and physicians have put their efforts for the exploration of "fuzzy for medical diagnosis" applicability. The study reveals that, fuzzy concept has been successfully implemented in the medical field. It has been widely accepted that the classical ontology techniques cannot sufficiently handle imprecise and vague knowledge for some real world applications, but fuzzy ontology can effectively resolve data and knowledge problems with uncertainty.[2] Fuzzy Ontology provides unique representation of knowledge. It enables unambiguous recording of data in the knowledge base. It is essential to capture signs and symptoms about diseases and store information about tests to be carried out to confirm a particular disease.

4. SYMPTOMS BASED DECISION ONTOLOGY

A classifier is an algorithm or method that assigns a class label to an object, based on the object description which may be structured or unstructured. It is also said that the classifier predicts the class label. The object description may be a vector containing values of the features (attributes) or structure which contain different field for producing the feature which deemed to be relevant for the classification task. Typically, the classifier learns to predict class labels using a training algorithm and a training data set. When a training data set is not available, a classifier can be designed from prior knowledge and expertise. Once trained, the classifier is ready for operation on unseen objects. [5] The various type of classifier is used in the data mining system. The K-NN classifier, SVM classifier, ANN based classifier, fuzzy classifier are the latest area of interest. There is various reasons to choose the fuzzy based classifier.

4.1 Soft Labeling

The standard assumption in pattern recognition is that the classes are mutually exclusive. This may not be the case, as the example in Figure 1 shows. A standard classifier will assign a single crisp label (rain). A fuzzy classifier can assign degrees of membership (soft labels) in all four classes {rain, clouds, wind, sunshine}, accounting for the possibility of winds and cloudy weather throughout the day. A standard classifier can output posterior probabilities, and offer soft labelling too. However, a probability of, say, 0.2 for cloudy weather means that there is 20% chance that tomorrow will be cloudy. A probabilistic model would also assume that the four classes form a full group, i.e., snow, blizzards or thunderstorms must be subsumed by one of the existing four classes. Soft labelling is free from this assumption. A fuzzy classifier, D , producing soft labels can be perceived as a function approximator $D:F \rightarrow [0,1]c$, where F is the feature space where the object descriptions live, and c is the number of classes. While tuning such a function approximator outside the classification scenario would be very difficult, fuzzy classifiers may provide a solution that is both intuitive and useful.

4.2 Interpretability

Automatic classification in most challenging applications such as medical diagnosis has been sidelined due to ethical, political or legal reasons, and mostly due to the black box philosophy underpinning classical pattern recognition. Fuzzy classifiers are often designed to be transparent, i.e., steps and logic statements leading to the class prediction are traceable and comprehensible.

4.3 Limited Data, Available Expertise

Examples include predicting and classification of rare diseases, oil depositions, terrorist activities, natural disasters. Fuzzy classifiers can be built using expert opinion, data or both.[6]

5. MALARIA

The Malaria is most common disease in India. The symptoms of malaria are many time confusing with the normal cold or viral fever .The accurate and timely diagnosis of malaria infection is essential to control and cure the disease. There were an estimated 190–311 million clinical episodes of malaria, and 708,000–1,003,000 deaths in 2008 [16]. It becomes the 5th cause of death from infectious diseases worldwide in low income countries. It is a leading cause of disease and death among children under five years and pregnant women. Malaria is a disease caused by the protozoan parasites of the genus Plasmodium.

Malaria parasites are transmitted from one person to another by the bite of a female anopheles mosquito. The female mosquito bites during dusk and dawn and needs a blood meal to feed her eggs. Male mosquitoes do not transmit malaria as they feed on plant juices and not blood. There are about 380 species of anopheles mosquito but only about 60 are able to transmit malaria. Like all mosquitoes, anopheles breed in water - hence accumulation of water favours the spread of the disease.[17] High Fever with vomiting, sweating, cold stage is the most constant symptom of malaria.

6. METHODOLOGY

The ontology systems are generated for malaria identification using fuzzy ontology with 3 numbers of features. The feature is as follows: 1. Fever 2. Cold stage and 3. Sweating. The most common data use in the ontology is based on the structure type data .The figures below is explain the system development:
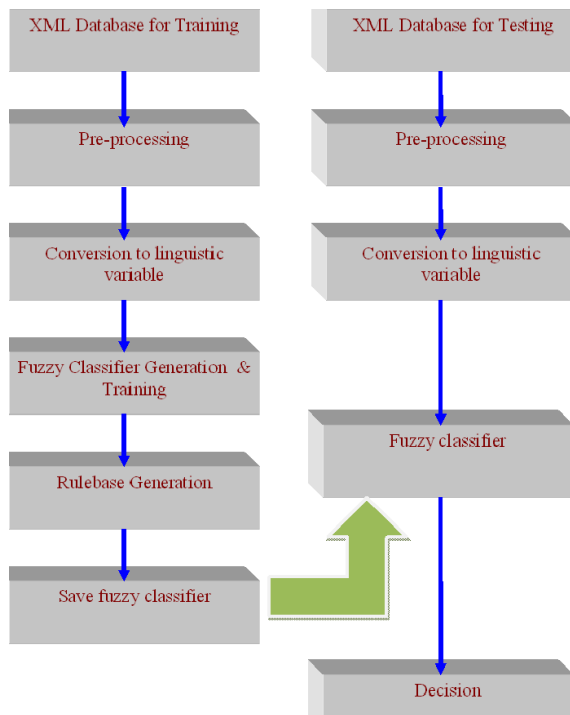
In this project xml based data input is used. Typical data file content is shown below
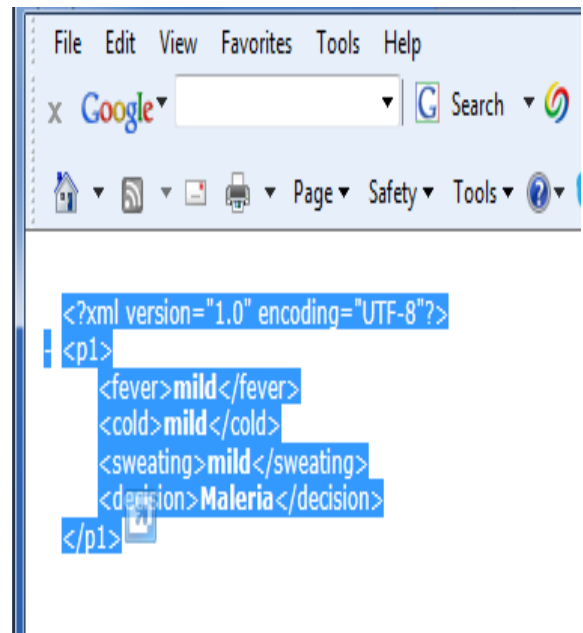

Fig. 1 Data input for the classifier system

The feature and decision are given in the training file. For testing file decision field is removed.

Data is read by the developed system and transfer to the classifier after processing. All The simplest fuzzy rule-based classifier is a fuzzy if-then system, similar to that used in fuzzy control. A fuzzy classifier is constructed by specifying clustering algorithm.

7. ALGORITHM FOR TRAINING OF THE SYSTEM

Step 1: For all N Domain Expert, an initial feature and decision are Loaded. Determine the numbers of linguistic variables associated with each input/output parameter.

Let D denotes the number of m tropical diseases, D (d1, d2, d3…..dm), where each di ∈ D , represent the set of all possible diseases.

Let S denotes the number of n symptom for a diseases,S (s1, s2, s3…..sn), where each si ∈ S, represent the set of all possible symptom.

Step 2: pre-processing of data. The s1…………… sn features is loaded and converted into numeric value for example for fever

Let Low represent the 0 value

Let mild represent the 1 value

Let high represent the 2 value

Step 3 : Implementation of clustering method for the fuzzy rule base construction and generation of fuzzy rules from the data.

Step 4: For all N Domain Expert, an initial FIS each is developed.

Step 5: optimize the number of cluster based on achieved decision accuracy on training set.

Step 6 : save the fuzzy classifier for decision generation.

8. ALGORITHM FOR TESTING OF THE SYSTEM

Step 1: For all N Domain Expert, an initial feature and decision are Loaded. Determine the numbers of linguistic variables associated with each input/output parameter.

Let D denotes the number of m tropical diseases, D (d1, d2, d3…..dm), where each di ∈ D , represent the set of all possible diseases.

Let S denotes the number of n symptom for a diseases,S (s1, s2, s3…..sn), where each si ∈ S, represent the set of all possible symptom.

Step 2: pre-processing of data. The s1…………… sn features is loaded and converted into numeric value for example for fever

Let Low represent the 0 value

Let mild represent the 1 value

Let high represent the 2 value

Step 3 : load the fuzzy classifier with associated rules for classification.

Step 4: apply the feature data to fuzzy classifier for decision generation.
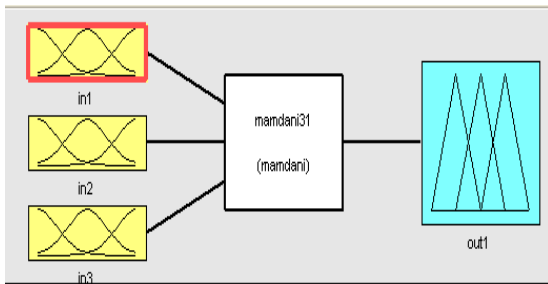
Fig. 2 FIS for ontology system

The three features in1,in2 and in3 are numerical but the rules use linguistic values. If there are M possible linguistic values for each feature, and n features in the problem, the number of possible different if-then rules of this conjunction type (AND) is Mn .In our case n is 6. Typical rule base for the proposed system is shown in figure 2.
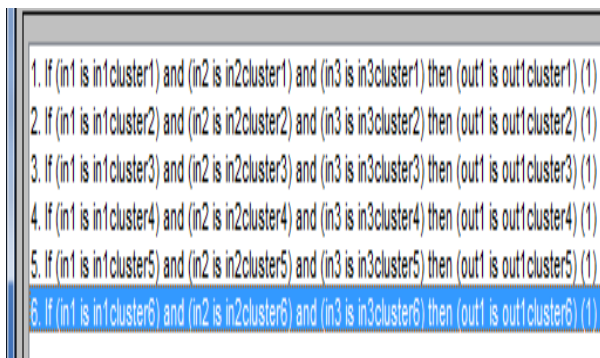

Fig. 3 Rule base of fuzzy classifier

If the fuzzy classifier comprises of all such rules, then it turns into a simple look-up table. Unlike look-up tables, however, fuzzy classifiers can provide outputs for combinations of linguistic values that are not included as one of the rules. Each linguistic value is represented by a membership function.
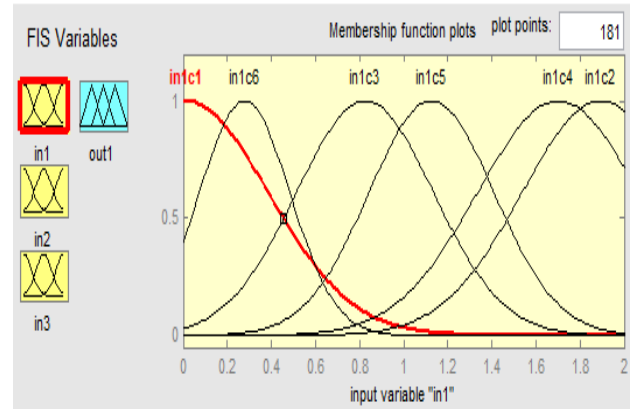

Fig. 4 Rule base of fuzzy classifier

9. RESULT

The Malaria diseases targeted in this ontology system. The MATLAB based software has been developed and tested. The following parameters are use for the simulation

Table I Parameters of simulation

S. No.	Specification	Value
1.	No. of feature	3 (Fever ,Cold and sweating)
2.	Classes of features	3 (low, mild and high) for fever and cols . 2 (Yes and No) for sweating
3.	Output class	2 (Malaria or No malaria)
4.	No. of instance	12
5.	Classifier	Fuzzy classifier based on clustering
6.	No. Of cluster for FIS	6

The feature and decision are given in the training file. Developed system reads the feature and decision from xml file and trains the FIS. The trained graph is shown below:
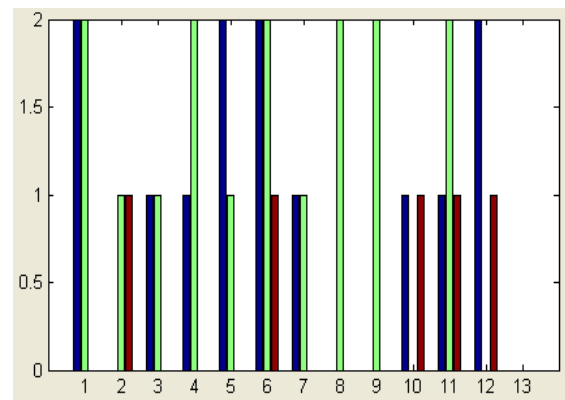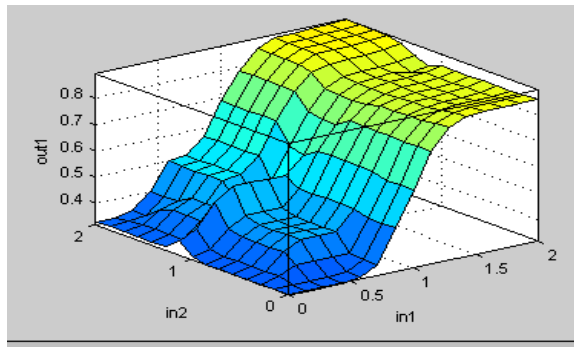

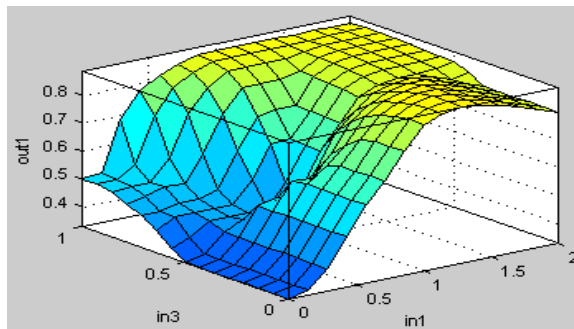**(a)**
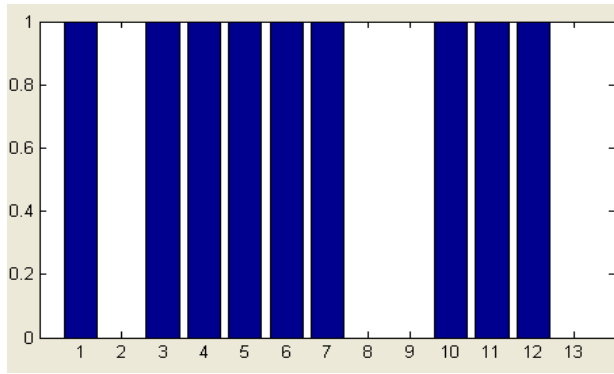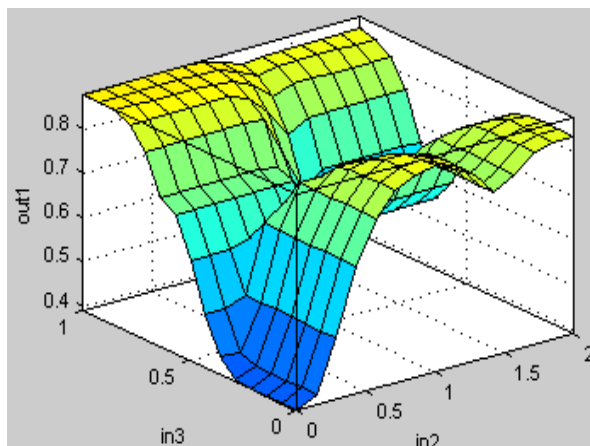
(b)

Fig.5 (a) Feature and (b) decision for the training of Fuzzy classifier
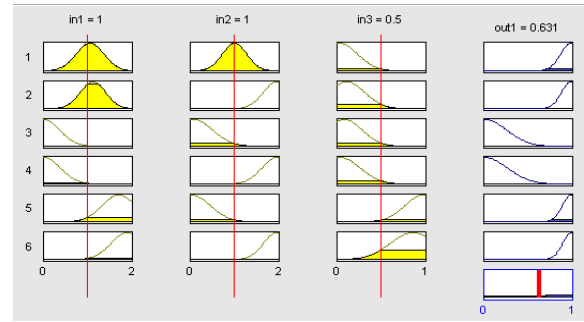

(a)

(b)

(c)

(d)

Fig. 6 Fuzzy classifier (a) Rule for 1st and 2nd input (b) Rules for 1st and 3rd input (c) Rules for 2nd and 3rd input (d) Result evolution from FIS

For testing file decision field is removed .Data is read by the developed system and transfer to the classifier after processing. The developed ontology system is giving 95 % accurate detection of malaria.

10. CONCLUSIONS

The fuzzy based ontology system is developed and tested on the database of 3 symptoms of Malaria disease. The fever , cold , sweating is used for malaria detection. These features are highly confusing for decision. The developed ontology system is giving 95 % accurate detection of malaria. In feature the accuracy become low when we add more systems and more disease. The optimization of rulebase only makes the reliable system.

11. ACKNOWLEDGEMENT

12. REFERENCES

[1] Fensel D., 2000. The semantic web and itslanguages. IEEE Computer Society 15, 6 (November/December), 67−73.

[2] Chang-Shing Lee, Senior Member, IEEE, and Mei-Hui Wang. A Fuzzy Expert System for Diabetes Decision Support Application, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 41, NO. 1, FEBRUARY 2011

[3] William Hsu, Member, IEEE, Ricky K. Taira, Suzie El-Saden, Hooshang Kangarloo, and Alex A. T. Bui, Member, IEEE.Context-Based Electronic Health Record: Toward Patient Specific Healthcare, IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, VOL. 16, NO. 2, MARCH 2012

[4] Barry Smith and Christopher Welty,Ontology: Towards a New Synthesis.

[5] http://www.scholarpedia.org/article/Fuzzy_classifiers

[6] Fuzzy Classifier Design, Springer-Verlag, Heidelberg, May 2000.

[7] "Russell S., Norvig P. *Artificial Intelligence: A Modern Approach.* Prentice Hall, Englewood Cliffs, NJ. (1995)

[8] Swartout B., Patil R., Knight K., Russ T. Toward distributed use of large−scale ontologies. In *Proceedings of the Tenth Knowledge Acquisition for Knowledge− Based Systems Workshop.* (KAW '96 November 9−14, Banff, Alberta, Canada). (1996)

[9] Noy N.F., McGuinnes D. L. 2001. ''Ontology Development 101: A Guide to Creating Your First Ontology''. Stanford Knowledge Systems Laboratory Technical Report KSL−01−05 and Stanford Medical Informatics Technical Report SMI−2001−0880, March.

[10] Fonseca, F. Egenhofer M., Agouris, P., Camara G. 2002. Using Ontologies for Integrated Geographic Information Systems. Transactions in GIS, −(6):3 in print.N. Lammari and E. Metais, "Building and maintaining ontologies: a set of algorithm," Data & Knowledge Engineering, vol. 48, no. 2, pp. 155-176, Feb. 2004.

[11] Runi Studer, V. Richard Benjamins, Dieter Fensel: Knowledge Engineering: Principles and Methods. Data Knowl. Eng. 25 (1-2): 161-197 (1998)

[12] V. W. Soo and C. Y. Lin, "Ontology-based information retrieval in a multi-agent system for digital library," The 6th conference on artificial intelligence and applications, pp. 241-246, Taiwan, 2001.

[13] C. S. Lee, Y. F. Kao, Y. H. Kuo and M. H. Wang, "Automated Ontology Construction for Unstructured Text Documents," Data & Knowledge Engineering, in press, 2006.

[14] H. C. Wang, C. S. Lee, and T. H. Ho, "Combining Subjective and Objective QoS Factors for Personalized Web Service Selection," Expert Systems with Applications, vol. 32, no. 2, 2007.

[15] X.Y. Djam and G.M. Wajiga, "A Novel Diagnostic Framework: The Applica-tion of Soft Computing Technology", The Pacific Journal of Science and Tech-nology, Volume 13. Number 1. May 2012 (Spring).

[16] World Health Organization (WHO), "World Malaria Report 2009,"Geneva: WHO Press, 2009, available at: http://www.who.int/malaria/world_malaria_report_2009/ en/index.html.

[17] Malaria. Greenwood BM, Bojang K, Whitty CJ, Targett GA. Review; Lancet 2005; 365:1487-98.

[18] https://semanticweb.com/why-ontologies-are-needed-in-health-care applications_b22669