A rule-based method for living organisms classification

A Desiani[1], M Arhami[2], Firdaus[3] and S I Maiyanti[4]

[1,4]Mathematics Departement, Faculty of Mathematics and Natural Science Universitas Sriwijaya,
[2]Informatics Engineering Department, Politeknik Negeri Lhokseumawe
[3]Faculty of Computer Science Universitas Sriwijaya

[1]anita.arhami@gmail.com, [2]Muhammad.arhami@gmail.com

Abstract. The implementation of an expert system for classification of a living organism could help researchers to classify an organism without having to face the complexities of a dichotomous key. This paper proposes a new system that can inference and give a solution for classification living organisms based on rule-based in taxonomy biology. The knowledge in taxonomy is represented by the rules-based method as a set know ledge for the system. the system is named as the expert system for living organisms. the system can be a solution to overcome obstacles in biological research caused by the problem of identification of living organisms. The system opens the possibilities to integrate other automated taxonomy systems because It can be used not only for a specific species but for all of other living organisms. The expert system of the classification is applied Peffer's DSRCM as a framework and guidelines. Forward chaining modified with backward chaining used as reasoning technique in the system. The expert system has tested on some groups respondent where all of the respondents give scores around 80-90 for the performance system. It shows that the system works well for classification living organims.

Keywords: Taxonomy, Classification, Living Organism, expert system, design search research

1. Introduction

An expert system is a concluded solution, derived from many solutions for solving a problem.The expert system is able to learn from some facts and decide right conclusion based on the given facts. An expert system is a part of the artificial intelligence that is able to make a conclusion, giving advice, explaining like a consultant, as well as having the ability to think logically[1]. An expert system (ES) is a application computer which has a special design to simulate an expert's abilities into the computer to solve a problem[2]. An expert system consists of 5 components[3], there are; 1) Graphical user interface, 2) Knowledgebase or learning base, 3) Database of facts, 4) Inference/reasoning engine, 5) Working memory(Figure 1). The expert system has the ability to capture, collect and compile knowledge from experts to an integrated system, sometimes some people name ES as Knowledge-based systems [4]. Expert systems have been extensively applied in various fields of research such as agriculture, medical, education, nanotechnology, control, scheduling, Geoscience, mineral identification, etc[2][3][5]–[12].

In the expert system, the technique of reasoning is a most crucial part of the system. Reasoning technique is a technique which is run by the engine so the machine can be an inference engine that has the ability to draw conclusions based on the facts. One of the familiar and famous reasoning technique is production rules (rule-base). production rules have been widely used in developing the expert system inference engine. Production rule has a simple form but powerful for knowledge representation. The production has flexible ability to combine declarative and procedural rules representation that is integrated into one form of a cognitive process's model of a set of rules [13].

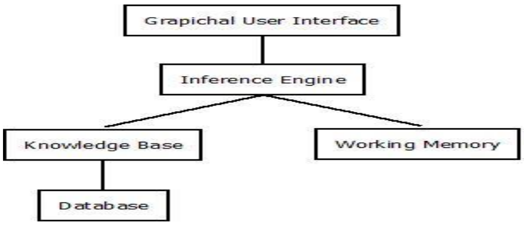

Figure 1. Expert System Components

Production rule or rule-based system can display human performance that the knowledge is seemingly got through experience[12]. Forward chaining and backward chaining are two methods of production rules for reasoning [12]. Forward chaining method (or bottom-up reasoning) is an inference that works from antecedents toward consequents. Otherwise, backward chaining begins work with a set of hypothesis or goals and performs to look for antecedents based on given consequents that support the consequents[14].

Classification is a technique in Taxonomy to classify living organisms as specified with their anatomical similarities, if the organisms have similar anatomical, they are put in one group.The organisms are grouped ranging from bigest organisms groups to smallest organisms groups. The orders of the groups are a kingdom(the greatest group), phylum (plural phyla), class, order, family (familia), genus (plural genera) and species (the smallest group)[15]. Classification is used for many applications. The first, the classification helps researchers to order organisms sequence. The Second, it is able to recognize newly discovered organisms based on the group that has similar with it. Third, it may ease the understanding process while learning the sorted organisms[16].

The method of organism classification is named as Taxonomy [16]. Taxonomy uses a dichotomous key for organism identification. A dichotomous key is a way of reasoning that provides some questions series to identify the organism that has only two choice answer yes/no. Every answer of the question will lead to next question until we get a conclusion about the organism. Reasoning in living organism needs precision and much time because living organisms in the world are very much and have a lot of variety[15]. .

The principle of classification process in Taxonomy has been used in several other filed such as software engineering (Design and Development)[17], Character recognition[18], fuzzy system and soft-computing[19][20], searching, and employee recruitment[21]. Learning application for the English Language[22]. In the field of Biology, taxonomy application has been developing on several specifics tasks, such as Dinoflagellates Classification using Neural network[23], species identification of fungi[24], automated insect identification system[25][26], Plant identification using image processing and computer vision[27][28], Snake identification[29].

The knock about identification of living organism that owns so many characters, becomes a major hindrance in biological research [30]. Reference [31], applied artificial neural networks to seek evolutionary relationships between living organisms but only for certain species., has developed machine to identify In the taxonomy, there are some disagreements about the kingdom system. There is a claim that in the world there is only 2 kingdoms system, but there are also other claims that there are 3 kingdoms system, or 6 kingdoms system, 7 kingdoms system, and 8 systems kingdom[32]–[36]. this paper will propose a new system for taxonomy that can inference the problem and give a solution based on rule-based in taxonomy biology. the knowledge in taxonomy biology is represented by the rules-based method as the knowledge for the system. the system is named as the expert system for living organisms. the system can be a solution to overcome obstacles in biological research caused by the problem of identification of living organisms. The system in this study will open the possibilities for other automated taxonomy systems because It can be used not only for a specific species but for all of other living organisms, although they are on the different kingdom systems.

Implementation of the expert system for the classification of living organisms can help people to classify an organism without having to face the complexities of a dichotomous key. The expert system classification of the living organisms also can be as a basic component to construct a larger system such

IOP Conf. Series: Materials Science and Engineering **403** (2018) 012082 doi:10.1088/1757-899X/403/1/012082

as eco-informatics. The expert system classification of the living organism will use a production rule as inference engine system and forward chaining as an inference method.

2. Method

This expert system of living organism classification is built by using Peffer's as based framework and guidelines. DSRCM has six phases that should go through from the identification of the problem phase until the communication phase[37]. Design-science research helps research paradigm pragmatic to find artifacts as real-life problem solving [38][39]. There are; concepts, models, methods, and instantiations[39][40]. DSRCM includes 6 phase:

1. Identity problem, this phase is to review the literature on the classification of living organisms and interview with various groups that have a direct relation to the classification of living organisms problem.
2. Establish the objectives for solutions. The Second phase is to identify the objects involved in the system to design solutions that are needed in the system. If the objects can be derived from the definition of the problem or the knowledge which obtained in phase 1. The selected object is an object that has a high rational in the problem. The results of this phase will be a guideline in designing an expert system for classification of living organisms.
3. Design and development. The third phase is built artifacts. In this paper, the obtained artifacts will become objects that have a contribution to the system. Some artifact will be a use case for the user in the system, a rational database, and knowledge inference models. in this phase, the system has just started to develop.
4. Demonstration. In the fourth phase, the artifacts that obtained in phase 3 is started to be implemented to develop an expert system for classification of living organisms. the system will be simulated and tested to see whether it is appropriate to user requirements or no and whether it provides solutions for users or no. This phase focuses on implementation of all the features in the design phase to build an expert system classification of living things using the programming language and database
5. Evaluation. In this phase, the artifacts in an expert system classification of living organisms will be observed and measured its performance if it can be completed and provide solutions to the problems that exist or no, whether all existing artifacts have features that correspond to the problem at an early stage. In addition, the user is prompted to interact directly with the expert system classification of living organisms to measure the performance of the system. This phase compares the solution of interest with the results obtained in the demonstration phase.
6. Communication. In this phase, the results obtained should be communicated between users, namely the public and researchers, to assist the further development system. The expert system for classification of living organisms can be integrated with a larger another system like information systems such as eco-informatics and bio-diversity system.

3. Result and Discussion

In taxonomy, classification of living organisms are divided into some levels depends on certain groups. the classification is arranged from the highest level of the kingdom to the lowest level of species. To facilitate the grouping of living organisms that are enormous and varied, so the classification will be grouped by some rules as in Figure 2.

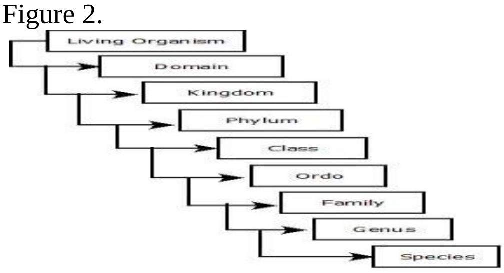

Figure 2. Taxonomy system

IOP Conf. Series: Materials Science and Engineering **403** (2018) 012082 doi:10.1088/1757-899X/403/1/012082

3.1. Database

From the process of data collection and interviews can be concluded that each of the characteristics of living organisms put them in different groups. The higher the level will have less in similar characteristics that must be considered, conversely the lower level of grouping, it will be more detail and more similar characteristics that must be considered. This classification system is useful for the identification of living organisms based on common characteristics in the taxonomy system. In the analysis phase, this phase obtains some Entities data and their relation to the system. It is explained by using the ERD (Entity Relationship Diagram) that can be seen in Figure 3.

The ERD will be implemented to database programming. The database is a basic component of knowledge in an expert system.

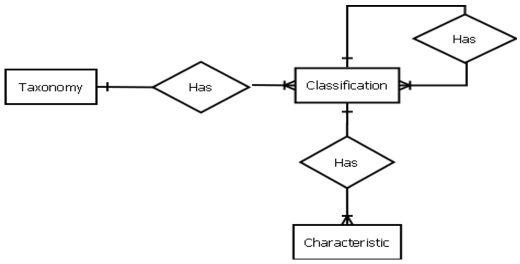

Figure 3. Entity Relationship Diagram

3.2. Knowledge Base

Designing knowledge has some several ways such as the semantic net, procedural representation, production rules, and frames[4]. For this research, the knowledgebase will be designed by production rule sometimes called IF-THEN Rule or Rule-based. Some paper used Production rule or Rule-base as a knowledgebase on their system. Reference [41] used production rule in an expert system for tourism in Jordan. Reference [42] applied production rule for nutrition care process of older adults expert system.

The premise of IF-Then Rules condition can combine attribute and logical expression like ‘or’ and ‘and’. When Clause IF is true, so the clause after THAN will be run. When the clause IF is false, so it should check the other rules. All of IF-Then rules in the classification of the living organism have two general form :

- IF FACT 1: class of taxon found in database AND FACT 2: character of organism is found in database THEN CONCLUSION 1: show name of organisms in the taxon
- IF FACT 3: Conclusion 1 AND FACT 4: new character of organism is found in database THEN CONCLUSION 2: show name of organism and its taxon

3.3. Inference Engine

The inference engine is The essential component in the expert system Inference Engine (IE) allows the ES to communicate the knowledge stored in the KB and to run the system, as it draws an inference by relating user-supplied facts to KB rules. This paper uses forward chaining that modified with backward chaining. Forward chaining is a bottom-up model where the solution to some problems naturally starts with the collection of information. Otherwise, backward chaining is a top-up model where is the model start from goal[43]. Inference uses this information to obtain logical conclusions for the solution[14]. Pseudo code for inferencing the classification of living organisms in the system is:

```
REPEAT
    FOR every rule do
            IF antecedents match assertions in the working memory and consequents would change
            the working memory THEN
            Create triggered rule instance
    END IF
    END FOR
Pick one triggered rule instance, using conflict resolution strategy if needed, and fire it (throw away
other instances)
UNTIL no change in working memory, or no STOP signal
```

IOP Conf. Series: Materials Science and Engineering **403** (2018) 012082 doi:10.1088/1757-899X/403/1/012082

For the example :
IF F1: taxon is class and F2: is Fishes **THEN**
(C1: Ordo is Primata OR C2: Ordo is Proboscidea) OR (C3: Ordo is Lagomorpha OR C4: Ordo is Marsupialia) OR (C5: Ordo is Rodentia OR C6: Ordo is Artiodactyla) OR (C7: Ordo is Dermoptera OR C8: Ordo is Cetacea) OR (C9: Ordo is Perissodactyla OR C10: Ordo is Insectivora) OR C11: Ordo is Carnivora

IF F3: ORDO is Carnivora AND (CF1: puny and cobby animals; AND CF2: cobby legs, globular ears, woolly) AND (CF3: deliver a strong-putrid secretion for sexual cueing and mark out area AND CF4: owns rectal whiff glands)
THEN C12: Familia is Mustelidae

IF F4: Familia is Mustelidae AND (FC4: its long body is between 173-217 mm or 6.8 to 8.5 inc AND FC5: the females are shorter than the males) AND (FC6: owns brown or red top layers and white maw AND FC6: some species of the population molt to a fully their white mantles in wintertime) AND (FC7: the population own lengthy and slim body, that permit them to adhere the prey into the hole AND FC8: They have tail with long are around 34-52 mm or 1.3-2.0 inc long)
THEN C13: Genus Mustela

IF F5: Genus Mustela AND (FC9: it has large and long tail with a protruding sooty tip AND FC10: its original from North America and Eurasia)
THEN C14: The Species of the organism is Stoat.

where, F_i : living organism class
FC_i :are Characteristics of living organism
C_i : Conclusion

3.4. Demonstration
The expert system was built to identify living organisms based on their taxonomy. Database implementation on the system used Mysql which still keep normalization rule and relational database. Expert system for classification is developed based on internet, so the system can be accessed anytime and anywhere. Expert system for classification of living organisms has some interface. Figure 4 is a user interface to input the level of taxonomy. This interface manages the level of taxonomy data for every kingdom systems. The interface of the system can provide two kingdoms system, five kingdoms system, etc. The other interface has a facility to manage characteristics of living organisms for every taxonomy level. This data will be used to identify organism based on their characteristic. This interface can see in Figure 5.

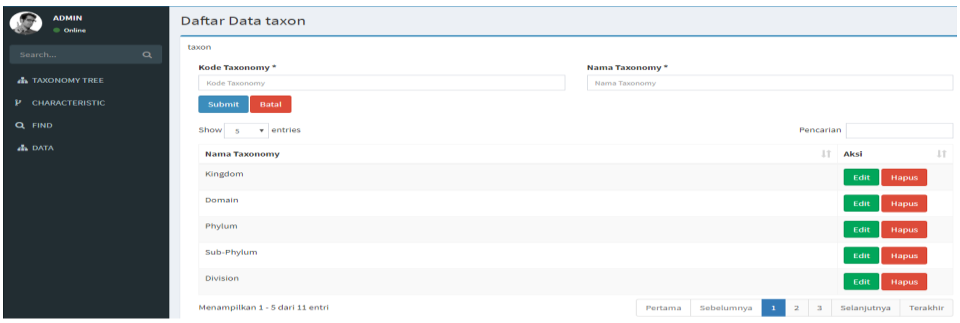

Figure 4. User Interface for Input Data Taxon

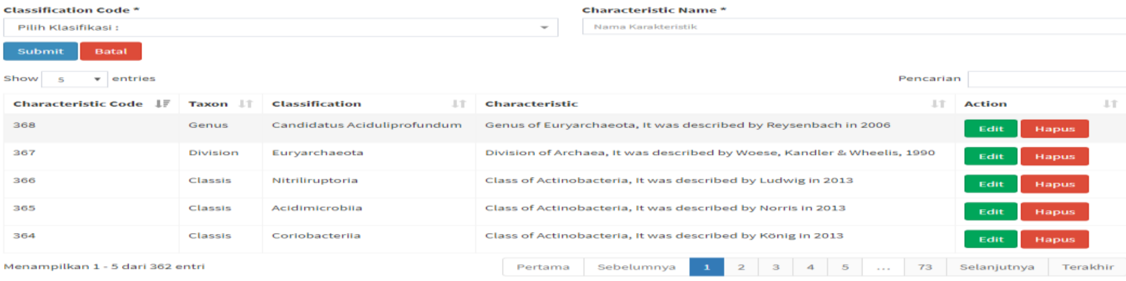

Figure 5. User Interface for Input Characteristic

For output, the system can draw the hierarchy classification tree of living organisms based on their group or based on their characteristic. So the user can look the three hierarchy of one various of living organisms or more (Figure 6) and also the user can see detail characteristic of organisms (Figure 7).

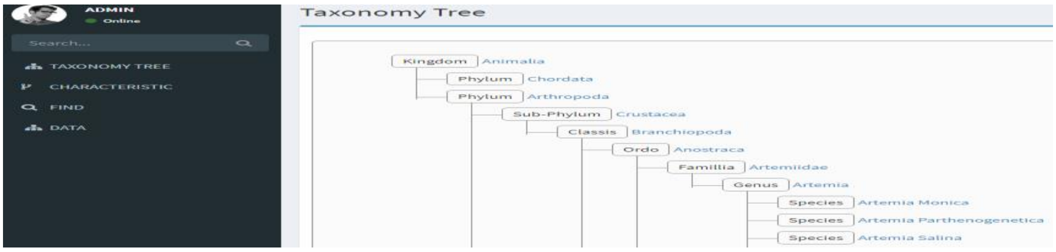

Figure 6. Taxonomy Tree Output Interface

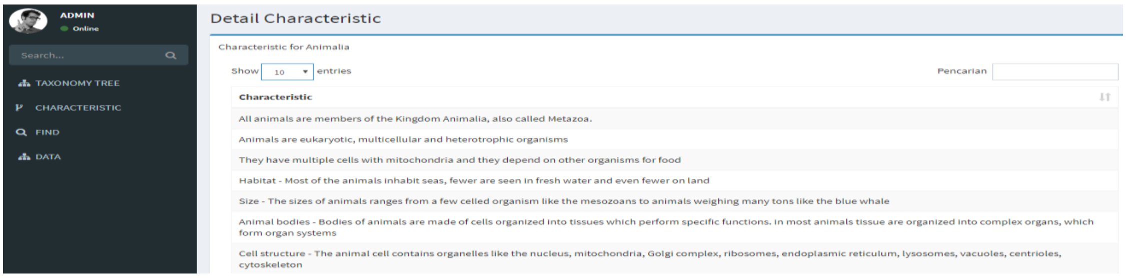

Figure 7. Characteristic Interface

4. Testing and Result

The Expert system application for classification of living organism has been testing on some groups respondent. The experiment is performed on 28 test subjects consist of 10 students from the different background, 10 undergraduate students from the biology department, 5 lecturers, and 3 researchers. The experiment subjects were tested with summated rating algorithms where the total rating is count. Summated rating is a quantitative method which the test subjects will give their response on favorable and unfavorable.

The result of the testing application gives the conclusion that expert system for classification of the living organism works well because all of the respondents give scores around 80-90 for the performance system. The respondents give suggestion the expert system for classification of living organisms can provide the better image with bigger size than existing image in the application.

5. Conclusion

The expert system application for living organisms used Peffer's DSRCM as a framework and guidelines to develop an application. The application can adopt every kingdom systems such as 2 kingdom 3 kingdom , and other system. Applied reasoning technique uses modified forward chaining with backward chaining because of the detail information in the classification of living organisms. The expert system for classification of the living organism is a quick, easy and precise solution to identify an organism without having to face the complexities of the dichotomous key.It has been shown by the respondents scores while they tested the application. An expert system classification of living organisms is a basic component to construct a larger system such as an ecoinformatics.

References

[1] E. Turban, *Decision Support System and Expert Systems*. USA: Prentice Hall International Inc, 1993.
[2] S. P. R. P. Isakki, "The Expert System Designed To Improve Customer Satisfication," *Adv. Comput. Int. J. ACIJ*, vol. 2, no. 6, pp. 69–84, 2011.
[3] K. . olorunso, I. O, Abikoye, O. C, 2 Jimoh, R. G. and 2 Raji, "A Rule-Based Expert System for Mineral Identification," *J. Emerg. Trends Comput. Inf. Sci.*, vol. 3, no. 2, pp. 205–210, 2012.
[4] S. Liao, "Expert system methodologies and applications — a decade review from 1995 to 2004," *Expert Syst. Appl. Elsevier*, vol. 28, no. June 2004, pp. 93–103, 2005.
[5] H. Ahmadi, M. Nilashi, L. Shahmoradi, and O. Ibrahim, "Computers in Human Behavior Hospital Information System adoption : Expert perspectives on an adoption framework for Malaysian public hospitals," *Comput. Hum. Behav.*, 2016.
[6] J. P. Astruc and R. Tufeu, "Knowledge Base Specification to Automate The Fluid," *Applied Artificial Intelligence*, no. scholarship 18860. pp. 453–471, 2001.
[7] R. E. Plant and M. P. Vayssie, "Combining expert system and GIS technology to implement a state-transition model of oak woodlands," *Elsevier Comput. Electron. Agric.*, vol. 27, pp. 71–93, 2000.
[8] B. Prasadl, P. E. S. N. K. Prasad, and Y. Sagar, "An Approach to Devepol Expert System In Medical Diagnosis Using Machine Learning Algorithms and A Perpormance Study," *Int. J. Soft Comput. IJSC*, vol. 2, no. 1, pp. 26–33, 2011.
[9] L. Salekhova, A. Nurgaliev, R. Zaripova, and N. Khakimullina, "The Principles of Designing an Expert System in Teaching Mathematics," *Univers. J. Educ. Res.*, vol. 1, no. 2, pp. 42–47, 2013.
[10] A. M. Salem, "Reasoning Techniques for Diabetics Expert Systems," *Procedia - Procedia Comput. Sci.*, vol. 65, pp. 813–820, 2015.
[11] G. W. Sasmito, "Application Expert System of Forward Chaining and The Rule Based Reasoning For Simulation Diagnose Pest and Disease Red Onion and Chili Plant," in *Proceedings of The 1st International Conference on Information Systems For Business Competitiveness (ICISBC)*, 2011, pp. 392–398.
[12] P. T. Sharma, P. N. Tiwari, and P. K. Shah, "Student Counseling System : A Rule-Based Expert System based on Certainty Factor and," *Int. J. Appl. Innov. Eng. Manag.*, vol. 2, no. 1, pp. 371–375, 2013.
[13] G. Giarattano, Joseph, Riley, *Expert Systems Principles and Programming*. Boston: PWS-KENT Publishing Company, 2005.
[14] D. K. P. Sharma, Tilotma, N. Tiwari, "Study of Difference Between Forward and Backward Chaining," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 2, no. 10, pp. 271–273, 2012.
[15] M. A. Ruggiero *et al.*, "A Higher Level Classification of All Living Organisms," *Plus One Open Access J.*, pp. 1–60, 2015.
[16] S. Rudyshyn and V. Samilyk, "Development Of Knowledge of The Taxonomy and Phylogeny Of Living Organisms for Future Biology Teachers," *Adv. Sci. J.*, vol. 2015, no. 1, pp. 75–82, 2015.
[17] M. Usman, R. Britto, J. Börstler, and E. Mendes, "Taxonomies in software engineering : A Systematic mapping study and a revised taxonomy development method," *Inf. Softw. Technol.*, vol. 85, pp. 43–59, 2017.
[18] T. Bouwmans, L. Maddalena, and A. Petrosino, "PT US CR," *Pattern Recognit. Lett.*, 2016.
[19] A. Saeed, S. Hafizah, A. Hamid, and M. B. Mustafa, "The experimental applications of search-based techniques for model-based testing: Taxonomy and systematic literature review," *Appl. Soft Comput. J.*, vol. 49, 2016.
[20] A. Fernández, V. López, M. José, and F. Herrera, "Knowledge-Based Systems Revisiting Evolutionary Fuzzy Systems : Taxonomy , applications , new trends and challenges," *J. Knowl.-Based Syst. KNOSYS*, vol. 80, pp. 109–121, 2015.
[21] G. Michel, "Levelized Taxonomy Approach for The Job Seeking / Recruitment Problem," in *IEEE International Conference on Embedded and Ubiquitous Computing, and International*

Symposium on Distributed Computing and Applications to Business, Engineering and Science, 2016, pp. 448–451.

[22] E. Yadegaridehkordi and M. Alaa, "Mobile Learning for English Language Acquisition : Taxonomy , Challenges , and Recommendations," *IEEE Access*, vol. 5, 2017.

[23] T. A. Meehan-Andrews, "Teaching mode efficiency and learning preferences of first year nursing students," *Nurse Educ. Today*, vol. 29, no. 1, pp. 24–32, 2009.

[24] M. Hsu *et al.*, "Species identification of medically important fungi by use of real-time LightCycler PCR," *J. Med. Microbiol.*, no. 2003, pp. 1071–1076, 2017.

[25] P. J. D. W. M. A. O. N. K. J. Gaston and I. D. Gauld, "Automating insect identification : exploring the limitations of a," *J. Appl. Entomol.*, vol. 123, no. 1, 1937.

[26] R. Gurgel-gon, E. Komp, L. P. Campbell, A. Khalighifar, H. L. Owens, and K. De, "Automated identification of insect vectors of Chagas disease in Brazil and Mexico : the Virtual Vector Lab," *PeerJ*, vol. 5, pp. 1–23, 2017.

[27] J. Joaci, D. M. Sá, P. César, R. Marta, and A. Ricardo, "images of palisade parenchyma," *Ecol. Inform.*, vol. 39, no. April, pp. 119–122, 2017.

[28] P. Barré, B. C. Stöver, K. F. Müller, and V. Steinhage, "A computer vision system for automatic plant species identi fi cation," *Ecol. Inform.*, vol. 40, no. May, pp. 50–56, 2017.

[29] V. H. K. R, MAHESHWARI. V, KUMAR V. K, "A Neural Network Based Specie Identification Technique in Venom Interacted Cases in India," *J Venom Anim Toxins Incl Trop Dis*, vol. 13, pp. 766–781, 2007.

[30] K. J. Gaston and M. A. O. Neill, "Automated species identification : why not?," *Philos. Trans. R. Soc. B Biol. Sci.*, vol. 359, no. March, pp. 655–667, 2004.

[31] A. Kasperski and R. Kasperska, "A new approach to the automatic identification of organism evolution using neural networks," *BioSystems*, vol. 142–143, pp. 32–42, 2016.

[32] T. Cavalier-smith, "Only six kingdoms of life," in *Proceedings of the Royal Society B: Biological Sciences*, 2004, no. May, p. 1251.

[33] O. R. Anderson *et al.*, "The New Higher Level Classification of Eukaryotes with Emphasis on the Taxonomy of Protists," *J. Eukaryot. Microbiol.*, vol. 52, no. 5, pp. 399–451, 2005.

[34] G. E. Woese, C. R., & Fox, "Phylogenetic structure of the prokaryotic domain: the primary kingdoms.," in *Proceedings of the National Academy of Sciences*, 1977, pp. 5088–5090.

[35] M. G. Simpson, "Plant systematics," in *Science and Technology Book*, Second., Academic Press is an imprint of Elsevier, 2010.

[36] T. S. R. Gayathri, *Biology (Zoology)*. Chennai, India: Tamil Nadu TextBook Corporation, 2003.

[37] K. Peffers, "A Design Science Research Methodology for Information Systems Research," *J. Manag. Inf. Syst.*, vol. 24, no. 3, pp. 45–78, 2007.

[38] S. Y. R. Esearch, B. A. R. Hevner, S. T. March, J. Park, and S. Ram, "D ESIGN S CIENCE IN I NFORMATION," *MIS Q.*, vol. 28, no. 1, pp. 75–105, 2004.

[39] P. Offermann, S. Blom, M. Schönherr, U. Bub, and D. T. Laboratories, "Artifact Types in Information Systems Design Science – A Literature Review," pp. 77–92, 2010.

[40] O. Wongso, Y. Rosmansyah, and Y. Bandung, "Gamification Framework Model , Based on Social," in *International Conference on Technology, Informatics, Management, Engineering & Environment*, 2014, pp. 10–14.

[41] N. S. Hussein and M. J. Aqel, "ESTJ : An Expert System for Tourism in Jordan," *Procedia - Procedia Comput. Sci.*, vol. 65, no. Iccmit, pp. 821–826, 2015.

[42] T. Cioara *et al.*, "Expert System for Nutrition Care Process of Older Adults," *Future Gener. Comput. Syst.*, 2017.

[43] A. Al-ajlan, "The Comparison between Forward and Backward Chaining," *Int. J. Mach. Learn. Comput.*, vol. 5, no. 2, 2015.