Automated Web-Bot Implementation using Machine Learning Techniques in eLearning Paradigm

Muhammad Munwar Iqbal[1], Muhammad Farhan[2], Yasir Saleem[3], Muhammad Aslam[4]

[1, 2, 3, 4] Department of Computer Sciences and Engineering, University of Engineering and Technology, Lahore
[2] Department of Computer Science, COMSATS Institute of Information Technology, Sahiwal
[1]Department of Computer Sciences and Engineering, University of Engineering and Technology, Taxila

ABSTRACT

E-Learning is very powerful tool and technique that is used to train and educates the people in these days. A lot of world ranking degree awarding universities are begun offer to deliver numerous courses lectures for high school grade education to degree level and even at post-graduation level through eLearning and distance learning paradigm. This research paper extensively concentrates supervised and unsupervised learning methods & strategies which are helpful for the e-Learning framework to automated answer of learners and student inquiries. Primary disadvantage of e-Learning paradigm is not in time answers of learner's questions, which decrease the learning curve of the student. The key demand in e-Learning environment is to dealing with Machine Learning classification can give higher to learning techniques but not fully automated with this paradigm. Training data set is used form to training the machine and then test data set is used to validate the this approach. This paper analyzes the different machine learning technique and proposed a solution from using these techniques.

KEYWORDS: Artificial Neural Networks (ANN), Supervised Machine Learning (SML), Machine Learning (ML), Support Vector Machine (SVM), Web bot.

1.   INTRODUCTION

Machine learning has to main approaches for the learning purposes, first is inductive machine learning and second deductive machine learning. Both techniques have the different data sets like training data set, target data set and test data set. The principal step is gathering the dataset. In the event that an essential master is accessible, then s/he could propose which fields (properties, gimmicks) are the most useful. If not, then the least complex technique is that of "beast energy," which means measuring everything accessible with the expectation that the right (educational, significant) gimmicks might be segregated. Be that as it may, a dataset gathered by the "animal energy" strategy is not specifically suitable for prompting. It holds much of the time commotion and missing gimmick qualities, and in this manner requires huge preprocessing [9].

The second step is the information planning and information preprocessing. Contingent upon the circumstances, scientists have various systems to look over to handle missing information [10]. (Hodge &Austin, 2004) have as of late presented a review of contemporary strategies for outlier (clamor) identification. These specialists have recognized the methods' focal points and disserves. Example choice is utilized to handle commotion as well as to adapt to the infeasibility of gaining from substantial datasets. Occurrence determination in these datasets is an improvement issue that endeavors to keep up the mining quality while minimizing the specimen size [8].

It diminishes information and empowers an information mining calculation to capacity and work adequately with expansive datasets. There is an assortment of strategies for inspecting occurrences from an extensive dataset [9]. Characteristic subset choice is the methodology of recognizing and evacuating whatever number unimportant and excess peculiarities as could be expected under the circumstances [11]. This lessens the dimensionality of the information and empowers information mining calculations

to work speedier and all the more adequately. The way that numerous gimmicks rely on upon each other regularly unduly impacts the precision of directed ML order models. This issue could be tended to by developing new gimmicks from the essential list of capabilities [12].

This strategy is called gimmick development/conversion. These recently created peculiarities may prompt the making of more succinct and exact classifiers. What's more, the finding of serious gimmicks helps better understandability of the handled classifier, and a superior understanding of the scholarly idea.
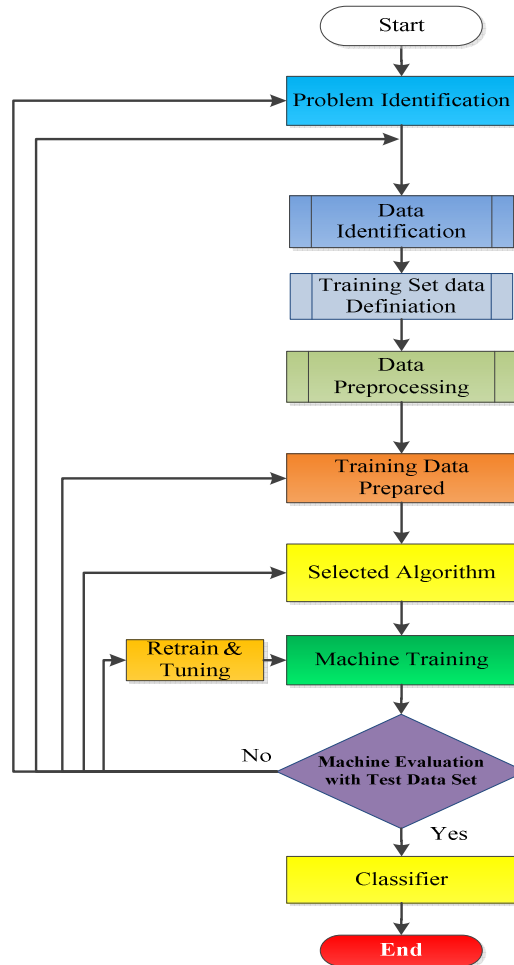

Figure 1: Supervised Machine Learning Work Flow

Some cases, where it is might be necessary to provide the pre-determined classifications to every single occurrence of a problem. If the moderator is so powerful which can grind out the groupings for him? This would stand a model of unsupervised learning in a classification perspective. Unsupervised Learning about the occurrences is unmarked. By smearing these unendorsed (clustering) algorithms, scholars have expectation to determine the unidentified, but valuable, types of entities [9].

2.   MOST RELATIVE PARADIGMS FOR ML

Machine learning in is an assorted field, detained together by basic objectives and comparable assessment techniques. The general point is to enhance execution on some job, and the general

methodology includes discovering and manipulating regularities in preparing data. Most assessment is investigational types, pointed at indicating that the learning in system prompts execution on a different test data sets, in one or more sensible areas, that is superior to execution on that test data set without learning. It is denoted that in-spite of these similitudes, scientists in machine learning have a tendency to link themselves with in the range of one five fundamental ideal models or paradigms, each model imparts essential suppositions about representation, execution strategies, and learning procedures.

2.1 Neural Networks
One significant standard is related with the field of neural network systems, speaks to learning as a multilayer system of units which manages to start from data input by inner elements to yield outputs. Evaluates on the connections figure out the amount enactment data is passed on. The initiations of yield outputs could be deciphered into numeric forecasts or discrete choices about the class of the data.
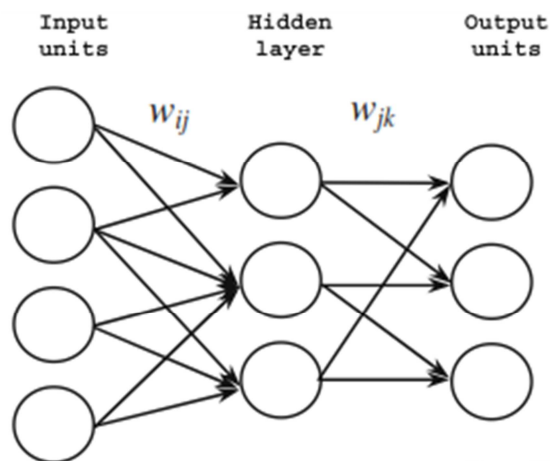
Fig. 2: Feed-forward ANN

Methodologies to adapting inside the neural net structure is commonly enhance order or forecasts correctness by changing the weights on the associations. One typical learning techniques in among the numerous that have been investigated, brings out ascent look through the weighted space by altering them, trying to minimize the failures which the system creates on preparing data set training. Rumelhart, Lehr and Widrow [15] sum up current research on neural network systems and portray a few provisions of this methodology.

2.2 Instance-based or Case-based Learning
A second structure or framework is called as case-based learning or instance-based learning, denotes to learning as expressions as particular cases or skills and depends on adaptable similar strategies to recover these cases and implement these to new circumstances. A typical plan which is known as nearest neighbor, basically discovers the already saved cases closest (as per few distance metric) to the recent circumstance and used for categorization or forecast. Case-based adaptive learning normally saves the formulating occurrences in memory; Uses of Machine Learning generalization happens at recovery time, a significant share in the indexing plan, the similarities metrics used to distinguish important examples (cases), however more modern variants may adjust a recovered case to the current states. Allen (1994) portrays this methodology, along some latest requisitions and implementation.

2.3 Genetic Algorithms
Third paradigm in machine learning is known as Genetic algorithms, commonly presents knowledge and information through Boolean or binary peculiarities, off and on again utilized as the conditions and

activities guidelines. A well-known translator for this learning utilizes an all-or-none matching procedure, utilizing qualities linked with guidelines to resolution of clashes.

In a few cases, creation framework structural design allow to runs discussed rules in grouping, transforming multi-step behavior. Learning administrators in Genetic algorithms, called hybrid and transformation in similarity to natural genetic systems which create new applicant principles from parents which possess best qualities, where each quality or "fitness" reveals the degree of efficiency on training cases. Basically, Genetic routines do parallel hill climbing which holding data struggling and frequently balancing depictions in memory. Goldberg [13] surveys genetic methodologies for the both optimization problems and machine learning.

2.4 Rule Induction

Rule induction is known as fourth paradigm, in which apply decision trees, condition-action rules, or comparable knowledge learning structures. In this paradigm, the efficiency component order cases down the extensions of the decision tree or discovers the first code whose conditions match the example, commonly utilizing an all-or-none comparing procedure.

Data about classes or forecasts are saved in the execution sides of the condition al rules and standards or the leaves of the tree. Learning procedure in the rule-induction skeleton generally help out a ravenous pursuit through the space of decision trees or tenet sets, ordinarily utilizing a factual assessment service to select traits for coping into the learning structure. Most techniques for the preparation data for information recursively into disjoint sets, endeavoring to outline every set as a conjunction of coherent rules. Quinlan [14] depicts a similar rule-induction procedure in some refined element.

2.5 Analytic Learning

A last method is in some cases labeled analytic learning that presents knowledge by means of guidelines in sensible structure and normally utilizes an efficient framework that settles multi-step issues utilizing some exploration methods. A typical method is to denote knowledge as Horn clauses as used in ROLOG and then to phrase issues as "hypotheses" and to scan for confirmations. Learning instruments in this paradigm by utilized basic knowledge develop confirmations or "clarifications" of experience, then aggregate the verifications into more multifaceted rules which can tackle comparative issues either with less inquiry utilizing neighborhood ""search-control rules" or in a solitary step by utilizing "macro-operators".

In analytical learning, mostly work has concentrated on enhancing the performance of search, however a few has managed enhancing precision on categorization jobs. The logics behind the different characters of these behaviors are more chronicled than investigative and scientific. The distinctive groups had their beginnings in diverse conventions and depend on diverse fundamental symbols.

In Fact an instance, exponents of neural network systems focuses correlations to neurobiology, case based learners to humanoid memory, researchers of genetic algorithm to advancement, authorities in rule induction inducing to heuristic search and supporters of analytic strategies to thinking in formal rationale. Someone can ask that this categorization profits the area, as dissimilarities of documentation and rhetoric frequently vague critical essential similarities.

3. PROBLEM STATEMENT

In eLearning paradigm learners and students confront the numerous issues that prompts the traditional structure is better education training framework to those individuals which can access it. Among these issues, late answers of the student inquiry are a real problem. In an eLearning environment learner is continually pushing for the right and in time replies. This issue prompts a hurdle for the learners in eLearning paradigm.

In eLearning paradigm, learner and educator cooperation is actually least level that is obstruction for the students to accept this paradigm. This problem could be resolved by implementing the diverse

machine learning techniques and methods. These approaches are exceptionally useful for eLearning nature's domains.

4. POSSIBLE SOLUTIONS

The identified problem can be solved with machine learning by making a chat-bot and training it with required material of the specific course and with semantic web; the responses from the semantic web system will be produced using a reasoner of the semantic web. The major difference between these two solutions will be the response; as the response of the chat-bot will be like human [1] and with semantic web reasoner the response will be document driven [figure 1].
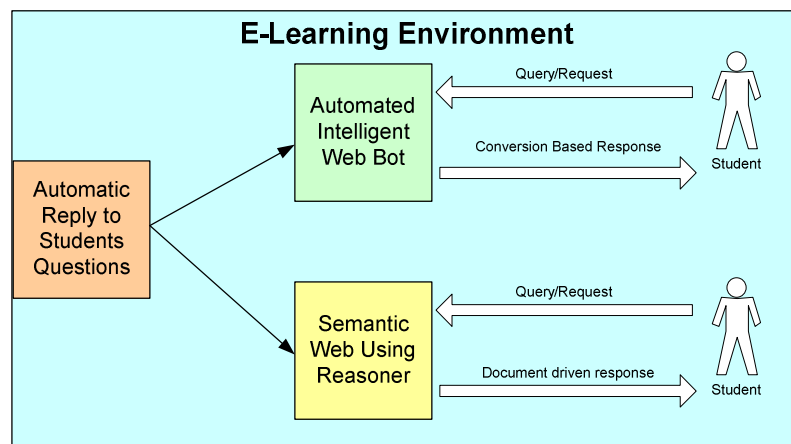

Figure 2: Possible solutions for automated reply to students' queries

Faqbot is a great thought to begin with search and application of the web-bot in eLearning paradigm. Chatter-bot for the execution of Turing test with $100,000 was a buzz and acquired win by the group of designers & developers working under the rules of Dr. Wallace for 3 times on yearly premise [2] [3] [4].Regularly made inquiries placed on the internet in stagnant structure don't give the learner a contentment which somebody is leading him, therefore it's a great idea to develop web-bot to reply queries and FAQs in a dynamic style[16]. Thusly the inquiry which are not recorded in the FAQs could be tended to future by adding answers with explanations for the bot.

Thus marketing or advertising group can work all the more successfully by knowing the issues of the clients as the logbook documents of bot give the data and information [1]. The need for the markup language emerges that is suggested by ALICE as Artificial Intelligence Markup Language (AIML) [5]. AIML is really straightforward and basic language that is focused around XML punctuation [1] [3]. (Atwell and Abu Shawar, 2004) has built a software for the transformation of plain content to AIML yet this software isn't accessible[6], therefore this conversion process requires to be completed , although the usage could be seen Pandora bots at World Wide Web.

5. MACHINE LEARNING APPROACH

Numerous machine learning methodologies have been portrayed in the research article, there is no technique that is best performs in all situations. Some comparative revisions are available, yet have not typically been performed sensibly and may lead to wrong results and conclusions. In (Table1) discussed the different approaches and their characteristics, which elaborates the suitability of the theses techniques for the machine in eLearning paradigm. This comparison gives a broad survey of studies identified with expert evaluation of advancement utilizing Machine-Learning Techniques (MLT).

Table 1: Machine Learning Techniques Suitability for the Proposed Solution

Sr. No.	ML Techniques / Parameters	Decision Tree Learning	Association Rule Learning	Artificial Neural Networks	Genetic Programming	Inductive Logic Programming	Support Vector Machines	Clustering	Bayesian Networks	Reinforcement Learning	Representation Learning	Sparse Dictionary Learning	Ontology Driven Semantic Web
1	Books, Handouts and lecture slides												
2	Plain Text as input		☐										
3	Structured input												
4	Algorithms	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
5	Ease of implementation	☐	☐	☐	☐	☐					☐		
6	Suitable for web Bot		☐	☐			☐						
7	Suitable for E-Learning												☐
8	Automated Tools Available	☐		☐				☐					☐
9	Text categorization	☐		☐	☐					☐	☐		

Machine adapting in current era is exhibiting the guarantee of generating reliably faultless appraisals. Machine learning framework adequately "realizes" how to gauge from training set of complete actions. The fundamental objective and commitment of the research is to help the exploration on expert estimation to studies utilizing machine learning methods.

This research exhibits the most typically implemented machine learning procedures; for example, case based reasoning, neural networks systems, rule induction, regression and classification trees and genetic algorithm. More over these machine learning techniques are show the comparative fulfilment of proposed solution of web bots. Techniques which are most suitable will be adopted for the automation of the query answer in electronic learning. The proposed solution will use the mining techniques as well as machine learning techniques to answer the learner queries, so for the generations of candidate solution machine learning techniques are used. The accuracy of these learning techniques is different in various paradigms.

6. PROPOSED SOLUTION

Components of the web-bot system are: NLP tools, Data Mining techniques, Machine learning technique, Web-bot, Reasonor, Knowledge Base and Teacher. Natural language processing deals with understanding, analyzing and producing the human context based language that people utilize regularly within request to interface with machines in both composed and spoken settings implementation characteristic human language rather than computer codes. One of the difficulties characteristic in natural language processing, to instructing computers to analyze it as the same fashion as people learn

and use languages. Over the span of human correspondence, the significance of the sentence relies on upon both the connection in which it was imparted and every person's understanding of the uncertainty in human language. This sentence postures issues for software which first modified to comprehend contexts and etymological structures. Natural language tool provides the facility to understand the semantic of the question. Three significant parts of any common language in understanding hypothesis: are Sentence structure, Semantics and Pragmatics.

The sentence structure depicts the manifestation of the language. It is normally indicated by syntax. Regular language is considerably more entangled than the formal language utilized for the manufactured language of rationales and workstation programs. The semantics gives the significance of the articulations or sentences of the language. Albeit general semantic hypotheses exist, when we fabricate a common language understanding framework for a specific provision, we attempt to utilize the least complex representation we can. Case in point, in the improvement that takes after, there is a settled mapping between words and ideas in the learning base, which is wrong for some spaces yet disentangles advancement. The practical segment clarifies how the articulations identify with the world. To comprehend language, an operator ought to consider more than the sentence; it need to consider the context of the sentence, the state of the world, the objectives of the speaker and the audience, uncommon meetings, and so on.

Data and Information Mining is a class of database applications which search for shrouded examples in group of collected information that could be utilized to forecast future behavior. Mining technique will be used to find the most relevant data for the solution of any question answer.

Machine learning technique will help the reasonor to producing the candidate solution and enhance the learning after every query. Machine learning includes streamlining a loss function on unlabeled information focuses given samples of marked information facts, where the loss function capacity measures the efficiency of learning algorithms. A review of methods that is called reductions for transformation of a problem reducing one loss function to problem of reducing an alternate, less difficult loss function.
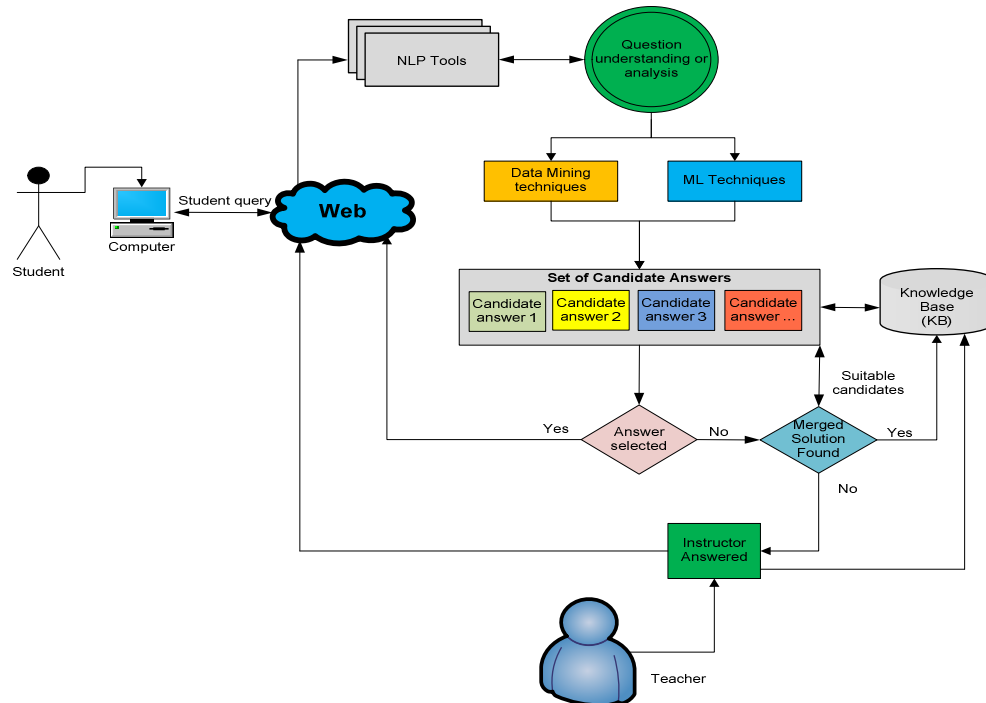

Figure 4: Automated Web-Bot Technique

Web-bot interact with student for the clear understanding of the student / learners question. The need for the markup language emerges that is suggested by ALICE as Artificial Intelligence Markup Language (AIML).

AIML is really straightforward and basic language that is focused around XML punctuation. Researcher has built a software for the transformation of plain content to AIML yet this software isn't accessible, therefore this conversion process requires to be completed, although the usage could be seen Pandora bots at World Wide Web. Programmed framework for utilizing Word Intelligent Handler is introduced along with 3 layered methodology.

$$p_o = \sum_1^N (p_i + p_{i-1})/2 \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (1)$$

Where i=1…………N and P0 denoted the coefficient of weighting, and this process starts with 1 to N where N is a natural number.

Reasonor provides the facility to answer the question on the basis of already understood semantics of the question. It can work only if and only if the data is well prepared and in process able format. Natural language annotation can also help the reasoner in providing such text. Annotation in the form of sentence structure and language linguistics. Enhancing the capabilities of the reasoner is also possible. This can be achieved by making the interaction of different reasoners in the particular scenario.

Knowledge base will help through already stored questions and answer of some specific queries and simultaneously for the new question answers. Knowledge base may contain the following components that help in find the answer of the query. It has most frequent questions of the student and answers with reference to there context. These question answers may be formatted in different Markup languages and called Knowledge Base Markup Language (KBML). It is somehow similar to HTML but special feature customized to satisfied the requirement of Knowledge Base. It also contains the full text search engine written in Perl or other language. A number of load balanced servers control the intelligent infrastructure paradigm.

The back end server's tiers handling the data, information and logics while presentation servers dealing with running Ubuntu Linux, Graphical User Interface and Web Interfaces, attached with numerous databases like MySQL, which contains the meta data and contents. Editing and revision history logs on all types of files and maintenance are controlled by Revision Control System (RCS). Extra utilities incorporate an apparatus to study the hunt strings really presented by clients, and apparatuses for adding and uprooting connections to other KB content and for evolving metadata. KB records are composed in KBML however deciphered into XML before being put away in the database and conveyed to the presentation framework. The KB additionally has web and order line apparatuses for concentrated pursuits, and an online synergistic workflow framework.

As the time passes the system will be more mature and capable because the candidate answers present in the knowledge base will also increase in such cases where no suitable answer is found. If suitable answer is not found in the knowledge base then the system has two possibilities. One of them is all candidate answers will be searched and a new candidate answer will be produced after merging the candidate answers so a new answer will be added to the knowledge base. If the question is too vague or it is not possible to make a new answer after merging the candidate answers then instructor has to intervene.

Here comes the role of the teacher. This interaction is possible in such cases where insufficient data is provided to the web-bot or the question is out of the domain or question is so vague. Such scenarios arise in the beginning of the process [figure 4]. Teacher can also be help full for the automated system as he or she put some pre-defined questions about the lectures which are currently delivered in eLearning paradigm

7. CONCLUSION FUTURE WORK

Our vision is that by using machine learning vast area of research through which the e-learning may be boosted up to limits. This research paper proposes an environment in through student can be get answers in a better and efficient manner. It is only possible if the web bot is working up to the extent

that needed to reply the student without any issue. There are number of machine learning techniques which are discussed with their pros and cons, so that produce an extraordinary solution of the student queries. In future we will try to solve the automatic answering by using web semantic techniques. Web-bot implementation is possible by preparing the dataset of annotated natural language text. Replying through web semantic web bot may results in better solution. Classic web provides millions of webpages e.g. Google search. By using our technique the students of the users will get only concise answers or search results. In future more machine learning algorithm and web semantic techniques will be implemented for the query and answer enhancement. It will very powerful too for the students, learners and also for the trainees.

REFERENCES

1. Abu Shawar Bayan, Atwell Eric and Roberts Andy.,FAQchat as an information retrieval system. In: Zygmunt V. (ed.), Human Language Technologies as a Challenge for Computer Science and Linguistics: Proceedings of the 2ndLanguage and Technology Conference, Wydawnictwo Poznanskie, Poznan, (2005), pp. 274-278.
2. Saygin, A. P., Cicekli, I., &Akman, V., Turing test: 50 years later. In The Turing Test, Springer Netherlands (2003), (pp. 23-78).
3. Abu Shawar, B. & Atwell, E., A chatbot system as a tool to animate a corpus. ICAME Journal, 29, (2005), 5–24.
4. Dr. Richard S. Wallace, ALICE AI Foundation retrieved on May 17, 2012, from, http://www.alicebot.org/bios/richardwallace.html
5. R. Wallace, The Elements of AIML Style, ALICE A.I. Foundation, 2003.
6. B. A. Shawar, E. Atwell, "A chatbot as a novel corpus visualization tool", Proceedings of Fourth International Conference on Language Resources and Evaluation (LREC'04). Vol.VI, pp. 2057-2060, 2004.
7. D. Hisgen& D. López De Luise., Dialog Structure Automatic Modeling. MICAI. ISBN 978-3-642-16772-0. Mexico, (2010).
8. Hodge, V., Austin, J., A Survey of Outlier Detection Methodologies, Artificial Intelligence Review, Volume 22, Issue 2, (2004), pp. 85-126.
9. Kotsiantis, S., Zaharakis, I., Pintelas, P.: Machine learning: a review of classification and combining techniques. Artificial Intelligence Review 26 (2006) 159–190
10. Batista GE, Monard MC., An analysis of four missing data treatment methods for supervised learning. ApplArtifIntell 17, (2003) :519–533
11. Lei Yu and Huan Liu. Feature selection for high-dimensional data: A fast correlation-based filter solution. In Proceedings of the Twentieth International Conference on Machine Learning (ICML 2003), 2003.
12. T.O. Ayodele, "Types of Machine Learning Algorithms". New Advances in Machine Learning, Yagang Zhang (Ed.) ISBN: 978-953-307-034-6. 2010.
13. Goldberg, D. E. Genetic and evolutionary algorithms come of age. Communications of the ACM, (1994). 37, 113-119.
14. Quinlan, J. R., C4. 5: programs for machine learning (Vol. 1). Morgan kaufmann. (1993).
15. Widrow, B., Rumelhart, D. E., k Lehr, M. A., Neural networks: Applications in industry, business, and science. Communications of the ACM, (1994). 37, 93-105.
16. Farhan, M.; Munwar, I.M.; ET. AL., "Automated Reply to Students' Queries in E-Learning Environment Using Web-BOT," Artificial Intelligence (MICAI), 2012 11th Mexican International Conference on , vol., no., pp.63-65, Oct. 27 2012-Nov. 4 2012 doi:10.1109/MICAI.2012.18 [URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6387216&isnumber=6387211]
17. AHMAD, Z., BABAR, M. R., & Ahmad, F., Verification of Identifier and Reserved word of Lexical Analyzer Using Z Notation. VFAST Transactions on Software Engineering, (2013). 1(1), 19-27.
18. JAWAD, S., & WAHAB, F., A Student-Centered Effective Learning Framework for Quality Education. VFAST Transactions on Education and Social Sciences, 3(1) (2014).