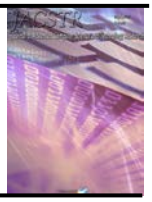
Design for Scientific RENAISSANCE



Exploring Qur'an by using Aspects and Dependencies

H. Hamam[1,2], M.T. Ben Othman[3], A. Kilani[1], F. Ncibi[1], M. Ben Ammar[1]

[1]Faculty of Engineering, Université de Moncton, NB, E1A 3E9, Canada
[2]Canadian Institute of Technology, Tirana, Albania
[3]Computer Science Dept., College of Computer, Qassim University, Kingdom of Saudi Arabia
Emails: [2]Habib.Hamam@ieee.org, [3]mtothman@gmail.com, [3]maathaman@qu.edu.sa

ISSN: 2231-8852

ABSTRACT

The availability of intelligent data mining tools are vital to help explore and comprehend the religious text of Islam. In this paper we present the research background of a platform offering an illustrative graphic-based decision aid tool enabling Qur'an experts to easily detect links between the multiple aspects presented in the Qur'an. This tool not only links one chapter to another chapter, or one verse to another verse through words, but also connects chapters and verses together through concepts and dependencies. As such, the platform is a self-evolving platform, interconnecting data through their aspects and dependencies. The user can expand the database by adding new dependencies for example. The platform is an extendible Web application. We propose combining several tools, mainly Java and PhP. This platform combines neural network in data mining and TFIDF algorithms to analyze and filter Qua'ran's content. The design of our approach is discussed and detailed.

***Keywords*: data mining, graphic-based decision aid, Qur'an multiple aspects, Qur'an dependencies**

1. Introduction

The Qur'an is believed to be the only holy document in history to preserve its content, structure, words, points, and accents conservatively. Muslims across the globe, regardless of their native language, should recite the Qur'an in Arabic, including during prayers. This aspect is extremely important when analyzing the Qur'an, as the set of words in Qur'an is fixed and widely memorized.

Each verse in the chapters (Sourats) of the Qur'an has a significant meaning; hence, the massive number of Qur'an interpretation books and studies. However, it is possible to minimize the number to the most famous or reputable ones. *Tanwir al-Miqbas min Tafsir Ibn Abbas*, " تنوير المقباس من تفسير بن عباس" , is an interpretation of the Holy Qur'an by Abd-Allah ibn Abbas [1]. Another prominent book is *Al Jam'e Al Ahkam Al Qur'an*, " الجامع لأحكام القرآن، والمبين لما تضمن من السنة وأحكام الفرقان"[2] by Al Qurtubi (1214-1273CE). This ten-volume book is a commentary on the Qur'anic verses dealing with legal issues. *Nukat al-Qur'an al-Dallah ala al-Bayan by Al Qassab* (d. 360AH/970CE) is a commentary discussing applied Islamic Law [3]. Other notable interpretation books include: [4] Abdul-Hameed al-Faraahee's book *At-takmeel fee usool at-ta'weel*; Shu'bah ibn al-Hajjaaj wrote, in the 8th centery, a book of interpretation, which was a collection of the opinions of the prophet's

companions and sahabah; Ibn Taymeeyah wrote a treatise called “Muqaddimah fee usool at-tafseer” in the 13th century; Ibn ‘Abdis-Salaam wrote a book about the Qur’anic metaphor; and in the 20th century, Shaykh Taahir al-Jazaa’iree wrote *At-tibyaan fee Uloom al-quraan*..

2. Problem statement

Data mining in the Qur’an is certainly necessary. For example, the relationship between monotheism (التوحيد), the worship of one God (عبادة الله), and the respectable and kind treatment towards one’s parents (برّ الوالدين), is exposed diversely and frequently in the Qur’an. The frameworks and verses discussing these themes can be dissected to explore these values’ relationships within the same Sourat, as well as their position within the order of revelation of the Sourates (ترتيب نزول السور) and verses. Of course, the context of revelation (أسباب النزول) must also be analyzed along with other elements. Readers can also question whether these links are isolated or interconnected with other issues, such as kinship (صلة الرحم), good neighborliness (وحسن الجوار), building a family, and raising children. Is there a link between all of the above mentioned aspects and goodness or corruption in the earth, as mentioned in chapters 2 (Sourat The Cow: سورة البقرة) and 13 (Sourat The Thunder: سورة الرعد)? The significance and relativity of the topics of the chapters (Sourats) combining the worship of God and kind treatment towards parents is studied as well to establish a broader context to these principles. Is there any relationship between them and the rules of Educational Psychology which, according to Western schools, does not include the worship of God as an important factor that may be part of the solution? etc. Data mining can expose the interlocking relationships of themes in the Qur’an’s to extract conclusions from which humankind can benefit. These conclusions may be rectified, formulated more precisely, or revised by the researcher who could introduce additional relevant data. The platform mining data should be flexible and user-friendly to be easily and efficiently used by researchers and experts.

Another example of the necessity of data mining in the Qur’an is the attempt to determine the human enemy who is obstructing an individual’s progress and the development of nations; and, who is threatening social peace as well as safe and decent living conditions. Could we find a coherent answer to this question in the Qur’an's Narration? According to Islamic beliefs, Satan is the greatest enemy of humankind, as explained in the in the Qur’an (chapter 35): “Indeed, Satan is an enemy to you; so take him as an enemy. He only invites his party to be among the companions of the Blaze.” Who is hampering the work of the prophets? To be more precise, who among human beings aids Satan against messengers? Are they the people (القوم) or an eminent group (الملأ) among the people? Who were these eminent groups throughout the eras? What kind of relationship exists between the people and the eminent group? What links all of these parties with the “Poets” (in this context, the Poets are those who are manipulate the people: الشعراء) and the topics discussed in the chapter of The Poets (سورة الشعراء)? What is the underlying meaning(s) of the verb “seduce” (أغوى), which is used to describe the actions of Satan and the Poets? What is the link between the prophets mentioned in the chapter of Poets and the poets as well as with the chapters mentioning the eminent groups? Who are the soldiers of Satan mentioned in the chapter of The Night Journey (الإسراء); and, what is their relationship with the horses and foot soldiers revealed in the same chapter? What are the relationships between the themes of the chapters, The Poets and The Night Journey? What is the relationship between the eminent groups and Haman (هامان)? What is the significance of the linking of the name “Haman” with “Pharaoh”, mentioned six times in the Qur’an in three different chapters? What are the links between the topics of these three chapters, and so on? Again, dissecting this web of themes can further provide some answers to these questions.

Data mining can certainly assist the reader in the analysis process. We offer a tool (platform or system) to expose complex relationships between the Qur’an’s data (its words and apparent meanings) for the benefit of the Qur’an reader, as well as that of the specialized scholar or

researcher. This tool should be malleable as to satisfy the requests of the users. .Scholars and researchers using this tool can guide the search by imposing new rules and/or by expanding the database. For instance, if we look for the relationships between data concerning seeing God on the Day After, there are texts in the Qur'an explicitly mentioning this event, "some faces, that Day, will be radiant. Looking at their Lord" (chapter 75), and "no! Indeed, from their Lord, that Day, they will be partitioned" (chapter 83). The user can enrich the database by adding, "for them who have done good is the best [reward] and extra" (chapter 10), wherein some scholars interpret the term "extra" as looking at the face of God. Concerning the aforementioned theme of kinship, it is discussed, although not explicitly, in three verses in the chapters Cow and Thunder (البقرة والرعد): "and sever that which Allah has ordered to be joined" (chapters 2 and 13) which contrasts the verse, "and those who join that which Allah has ordered to be joined". This information can be added to the flexible platform.

3. Related work

We have found few projects, surrounding text mining and statistical classification, which include a comprehensive study of data mining in the Qur'an.

Text Mining in Qur'an: In work [5], the author attempted to offer computer assisted Qur'anic text mining. He developed some online applications, such as word co-occurrence, chapter relatedness, Qur'an concordance, and verse similarity applications; he also developed two data sources—verse relatedness data source and pronoun reference in the Qur'an data source.

Statistical Classification: A statistical classifier of the Holy Qur'an Verses (Fatiha and Yaseen Chapters) has been proposed, in which a system has been conceived to categorize the verses in each chapter [6].

Generation of Frequent Patterns: Another method [7,8] consists in representing the Qur'anic text corpus as a graph, and applying a frequent sub-path mining algorithm on it to generate frequent patterns. The aim is to cluster and index similar verses of the Qur'an.

Discovering thematic interrelationships: Another method [9] consists in discovering thematic interrelationship between different Sourates (chapters) using their lexical frequency profiles. It has been applied successfully on 24 Sourates with more than one thousand words each.

Qur'anic Knowledge Map: Reference [10] suggested developing of a comprehensive Qur'anic Knowledge Map to provide on-line applications, including among others: morphological searching, interlinear translation, word sense disambiguation, and concept topic map.

Educational System for the Holy Qur'an and Its Sciences for people with special need: In a previous work we suggested a system navigating Qur'an to offer services to blind and handicapped people [11, 12].

4. Methodology

We present in this section the methodology used in our application. The neural network in datamining and TF-IDF algorithms are applied to analyze and filter Qur'an's content.

Advancing a Model of the Qur'an's Content: The model of Figure 2 is used for dissecting the Qur'an's content. The context of revelation may be added as a dependency. The output may be a graphic representation illustrating the network of a given aspect with its dependencies.

Aspects and Dependencies: The complexity of the system increases with the number of concepts and dependencies included. The database should adapt to this increasing complexity. The speed of the information processing will also depend on this complexity and should therefore, be optimized.

Hyperlink Data: The platform should also allow the hyperlinking of words and sentences in the Qur'an through the network of aspects and dependencies, as to ensure that the navigation of the data will be easily accessible and the concluding process more precise.

Self-development of the Platform: The proposed platform is self-evolving. A user, especially an expert, may enrich the database and contribute to its development. New dependencies and even new aspects with their dependencies may be added. After being review by other experts, they may be integrated in the platform.

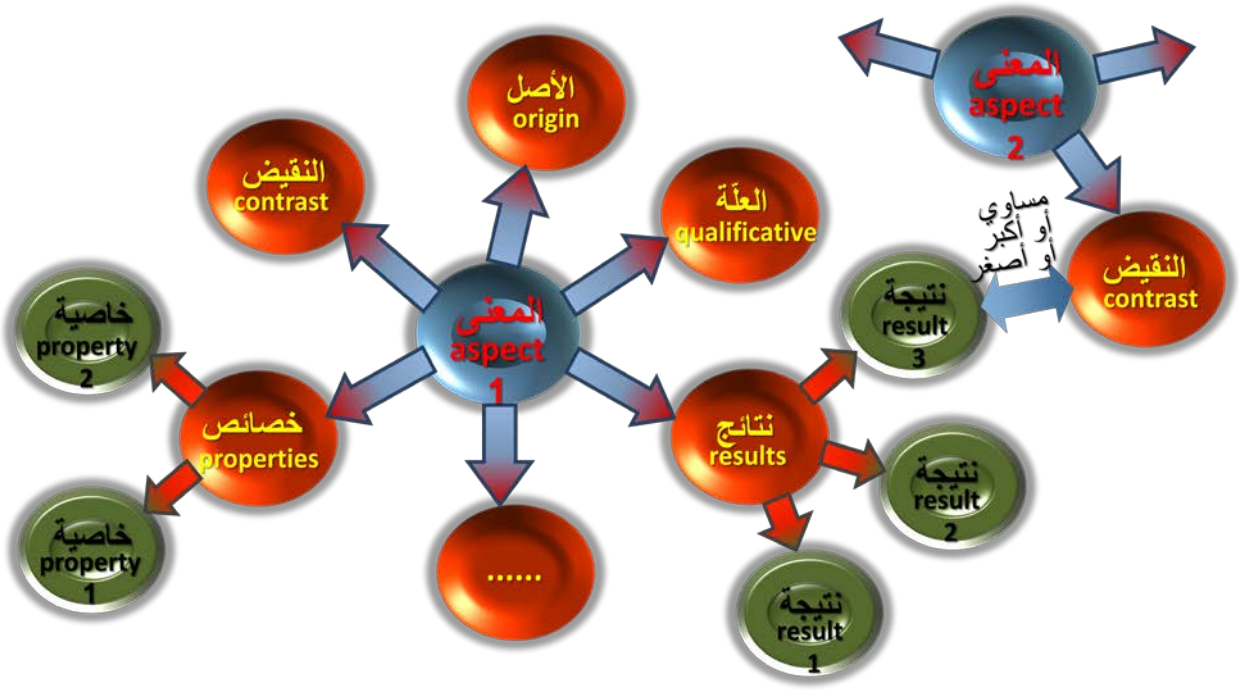


Figure 2: interconnecting data through dependencies data dependencies such as the origin (الأصل) the qualificative (العلة), the contrast (النقيض) the properties (chapter, position of the versus in the chapter, mekkite or medinite, specific for some people, etc...) and the results.

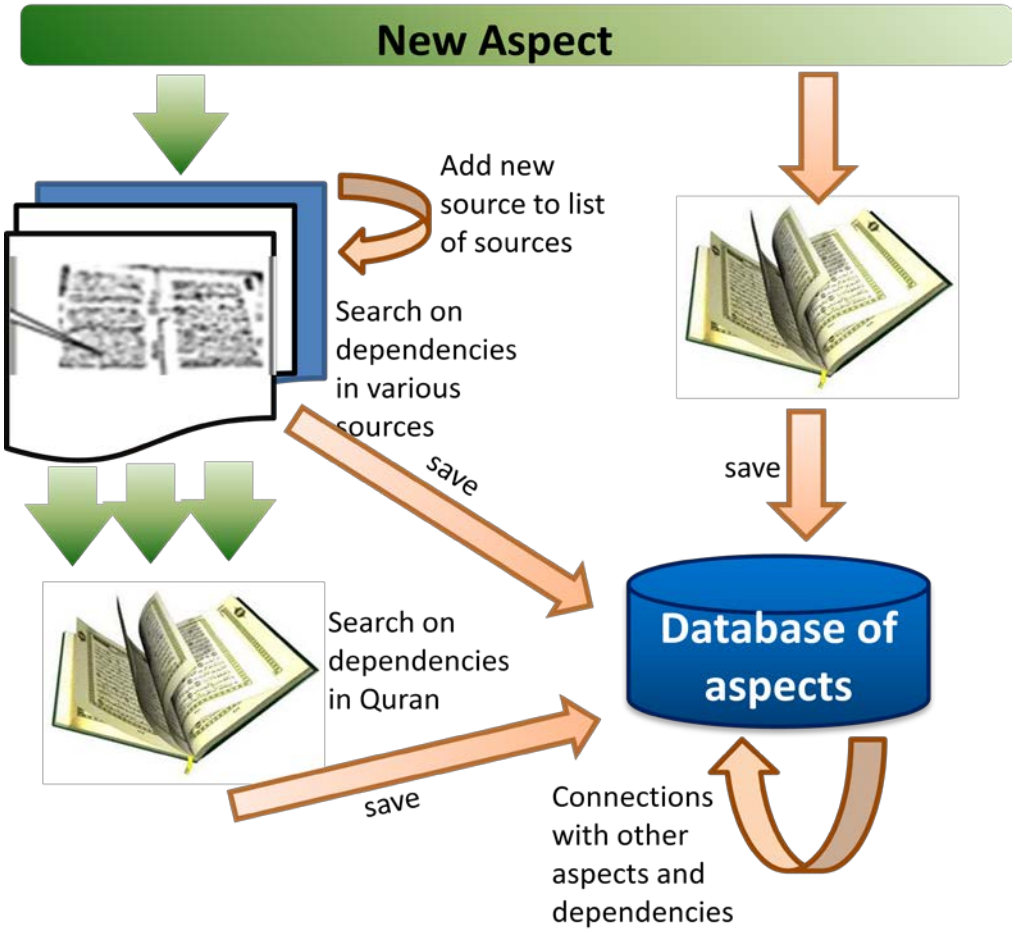


Figure 3 : Flowchart for entering a new Qur'anic aspects in the platform.

4.1 Neural Network in Data mining

Neural networks are used to model complex relationships between inputs and outputs or to find patterns in data [13-18]. This process is known as data mining [17]. The difference between these information databases and ordinary information databases is that there is a continual manipulation and reclassification of the data helping those who are seeking answers from the Quran to more easily extract the relevant & particular data that they need. The ability to do pattern recognition and function estimation is what makes artificial neural networks an efficient approach in data mining [17].

We can classify data mining tasks into two categories: descriptive and predictive data mining. Descriptive data mining can provide us with data so that we are able to understand what is happening without being given a pre-set idea or conclusion. Predictive data mining permits us to enter records that have unknown field values and it will try to forecast based on previous patterns.

The process of data mining is set up in the following stages: selection, pattern recognition, deployment.

Selection - This first stage is when the data is chosen (determine if useful or not), prepared, transformed and set up in order to move on to the next stage.

Pattern Recognition - This stage is where the construction of the models are built. We will choose the best model based on their best and most stable performances. This stage can be an elaborate and long process.

Deployment - This final stage is when the best model selected in the previous stage is now applied to the new data so that it can generate predictions on the best possible outcome.

Some research [13, 15, 17, 18] shows that neural networks performed better than conventional statistical approaches in many fields and are an excellent data mining tool. Ms Smita. Nirkhi [13], presented the potential use of artificial Neural Network in Data Mining.

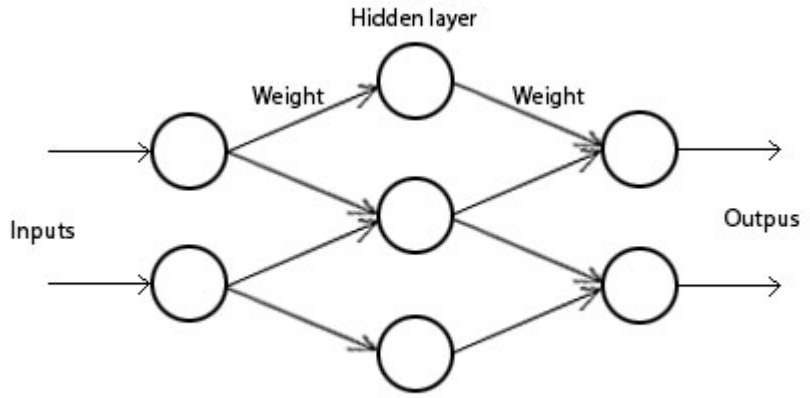


Figure 4: Neural network architecture.

4.2 TF-IDF Algorithm

TF-IDF (Term Frequency-Inverse Document Frequency) is a numerical/mathematical approximation that reflects the importance of a particular word in a given piece of text or document in a collection or corpus. TF-IDF is used as a weighting factor and its value increases depending on the number of times a particular word appears in the text or document. The value varies depending on the amount of times the word appears in the corpus. TF-IDF is widely used in data mining applications such as search engines, Web content exploration…. etc. The variations of the TF-IDF weighting scheme are frequently used by the search engines as a tool in ranking a text or a documents importance.

Several works were recently conducted to improve the TF-IDF algorithm [19-22]. Li and Guoyong [21] claimed that the TFIDF is the most adopted approach due to its efficiency and simplicity.

5. Algorithm & experimentation

We used multi-layered perceptron with back-propagation learning algorithms for both classification and regression type problems.

While most of Artificial Neural Network (ANN) algorithms start by using random weights, we proceeded differently in our experimentation (Fig. 4). Since we are using semantic similarity, the weights were calculated based on the ranking of the data mining map that we created.

The ‘data mining map' is the scheme that we used to describe the classification technique of our data in our data base.

Our data mining map is set up and classified as follows: by keywords, synonyms, categories, location, dates & time, sources, authors. Each time we enter a new aspect we build a new data mining map. The algorithm for entering a New Aspect is composed of the steps pointed out in Figure 3. Seven steps are accomplished when a new aspect is introduced:

1. Propose a new aspect,

2. Search in the Qur'an this aspect and identify where it is discussed or mentioned. The search should be intelligent since the wording of the aspect may not be explicit or direct. Keywords and an algorithm of similarity determination should be used. We may also search for similarities using existing sources, as some Qur'an experts have already established connections between similar Qur'anic aspects.

3. Concerning the proposed aspect, the values of the dependencies (time, location, sources) are searched through reliable sources and added into the database.

4. Any new reliable source found will be added to the list of sources.

5. Dependencies are then searched within the Qur'an and saved in the database.

6. Connections between the new entered aspects (and dependencies) and the existing content of the database are built.

7. After several iterations and connections our data mining map will be then enriched with new keywords and will continue to expand and build the potential relationships between a new aspect and the sourats. Through the iterations and comparisons, the ranking of keywords (sources, authors, categories) will continuously changed and improved.

The platform is an extendible Web application. We proposed combining several tools, mainly Java and PhP.

6. Results

The models are now designed. The kernel of platform has been conceived by using the framework codeigniter. We have also prepared a list of reliable sources to collect data. The process is certainly lengthy and requires human resources to enter a large number of data. We started by validating the concept with a few set of aspects.

The feed-forward for the back-propagation topology is very important, and there is no perfect solution on how to optimize the layout of the network. Each particular application has its own criteria and challenges.

To improve our results, we first tried to focus on the optimal selection of the parameters for the feed-forward for the back-propagation. Since the number of layers and the number of the processing elements per layer are big, we focused on our data mining map to better describe and understand the new sources and the 'sourats'.

With our choice of data mining map items, we were able to train and learn our back-propagation algorithm, and therefore to be able to filter relevant relationship between aspects and the Qur'an content.

7. Conclusion

Our objective is to create a platform offering an illustrative graphic-based decision aid tool enabling the users to data mine in Qur'an. In comparison with the existing applications mentioned above, our platform aims to benefit users far more. The model of aspects and dependencies gives us powerful tools to explore data in depth, as it analyzes links more profoundly. Our tool extends beyond linking one chapter to another chapter, or one verse to another verse through textual repetition and similarities. It connects chapters and verses together through concepts and dependencies, which presents a large interconnected net.

In the future, we propose bridging with hadiths. This is an immense task, of which detailing the process thoroughly goes beyond the scope of this work. For now, the platform will limit its attention to connections between Qur'anic aspect and hadiths using a limited set of keywords. Our purpose is to demonstrate this feasibility and prepare the ground for our future work.

Acknowledgment

This work was and is still supported by the Deanship of Scientific Research at Qassim University and the Canadian Institute of Technology.

References

[1] Tanwir al-Miqbas min Tafsir Ibn Abbas "تنوير المقباس من تفسير بن عباس", first edition, 1992, http://ia700807.us.archive.org/10/items/waq4125/4125.pdf

[2] Tafseer Al Qurtubi (Al Jam'e Al Ahkam Al Qur'an), "الجامع لأحكام القرآن", https://archive.org/details/TafseerEQurtubiArabicalJameAlAhkamAlQur'an

[3] Ahmad Al-Saiid Zaki Hemeidah, Repentance as a Legal Concept, pg. 26. Master's thesis for the Universoty Of Arizona's Department of Near Eastern Studies, 2011

[4] Usool at-Tafseer by Bilal Philips, Qur'anic Tafsir and Methods of Tafsir, http://www.iqrasense.com/Qur'an/Qur'anic-tafsir-and-methods-of-tafsir.html

[5] Abdul Baquee Muhammad, Online Application, 2012, http://www.textminingtheQur'an.com/wp-content/uploads/2013/11/firstYearTransferReport.pdf

[6] M.N. Al-Kabi et al. "statistical classifier of the Holy Qur'an Verses (Fatiha and Yaseen Chapters)", Asian Network for scientific Information, ISSN 1812-5654, 2005 http://198.170.104.138/jas/2005/580-583.pdf

[7] I. Ali "Application of a Mining Algorithm to Finding Frequent Patterns in a Text Corpus: A Case Study of the Arabic" International Journal of Software Engineering and Its Applications Vol. 6, No. 3, July, 2012, http://www.sersc.org/journals/IJSEIA/vol6_no3_2012/16.pdf

[8] Abdul Baquee Muhammad, Online Application, 2012, http://www.textminingtheQur'an.com/wp-content/uploads/2013/11/firstYearTransferReport.pdf

[9] N. Thabet, "Understanding the thematic structure of the qur'an: An exploratory multivariate approach", In Meeting of the Association for Computational Linguistics, (2005).

[10] E. S. Atwell, C. Brierley, K. Dukes, M. Sawalha and A. Sharaf, "An artificial intelligence approach to arabic and islamic content on the internet", In Proceedings of NITS 3rd National Information Technology Symposium, (2011).

[11] S. Mohamed, A. Hassanin, M Ben Othman "Virtual Learning System (Miqra'ah) for Qur'an Recitations for Sighted and Blind Students", JSEA 7, 195-205, 2014

[12] S. Mohamed, A. Hassanin and M. Othman "Educational System for the Holy Qur'an and Its Sciences for Blind and Handicapped People Based on Google Speech API". Journal of Software Engineering and Applications 7, 150-161. 2014

[13] Nirkhi, S., "Potential use of Artificial Neural Network in Data Mining," in *Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference on*, vol.2, no., pp.339-343, 26-28 Feb. 2010 doi: 10.1109/ICCAE.2010.5451537

[14] Sinha, N.K.; Gupta, M.M.; Rao, D.H., "Dynamic neural networks: an overview," in *Industrial Technology 2000. Proceedings of IEEE International Conference on*, vol.1, no., pp.491-496 vol.2, 19-22 Jan. 2000 doi: 10.1109/ICIT.2000.85420

[15] Xinjian Qiang; Guojian Cheng; Zheng Wang, "An overview of some classical Growing Neural Networks and new developments," in Education Technology and Computer (ICETC), 2010 2nd International Conference on, vol.3, no., pp. V3-351-V3-355, 22-24 June 2010, doi: 10.1109/ICETC.2010.5529527

[16] Yue-Seng Goh; Eng-Chong Tan, "An integrated approach to improving back-propagation neural networks," in TENCON '94. IEEE Region 10's Ninth Annual International Conference. Theme: Frontiers of Computer Technology. Proceedings of 1994, vol., no., pp.801-804 vol.2, 22-26 Aug 1994, doi: 10.1109/TENCON.1994.369201

[17] Dr. Yashpal, S.; Alock S. CH. "Neural Networks in Data Mining" Journal of Theoretical and Applied Information Technology, 2005 - 2009 JATIT.

[18] Cornelis Pieters "Into Complexity: A Pattern-oriented Approach to Stakeholder Communications", Universal-Publishers, 2010, ISBN 1599423057, 9781599423050

[19] Shouning Qu; Sujuan Wang; Yan Zou, "Improvement of Text Feature Selection Method Based on TFIDF," in Future Information Technology and Management Engineering, 2008. FITME '08. International Seminar on , vol., no., pp.79-81, 20-20 Nov. 2008 doi: 10.1109/FITME.2008.25

[20] Tian Xia, Yanmei Chai, "An Improvement to TF-IDF: Term Distribution based Term Weight Algorithm" Journal of software, vol. 6, no. 3, march 2011. © 2011 ACADEMY PUBLISHER doi:10.4304/jsw.6.3.413-420

[21] Bin Li, Yuan Guoyong, "Improvement of TF-IDF Algorithm Based on Hadoop Framework" The 2nd International Conference on Computer Application and System Modeling (2012).

[22] Li-Ping Jing, Hou-Kuan Huang, Hong-bo Shi "Improved feature selection approach TFIDF in text mining" Proceedings of the First International Conference on Machine Learning and Cybernetics, Bejing, 4-5 November 2002, 0-76803-7508-4/02/$17.00 2002 IEEE.