# Global Journal of Advanced Computer Science and Technology

**Volume No. 11**

**Issue No. 2**

**May - August 2023**

**ENRICHED PUBLICATIONS PVT. LTD**

**S-9, IInd FLOOR, MLU POCKET,**
**MANISH ABHINAV PLAZA-II, ABOVE FEDERAL BANK,**
**PLOT NO-5, SECTOR-5, DWARKA, NEW DELHI, INDIA-110075,**
**PHONE: - + (91)-(11)-47026006**

# Global Journal of Advanced Computer Science and Technology

**Aims and Scope**

Global Journal of Advanced Computer Science & Technology deals with information technology, its evolution and future prospects and its relationship with the Business Management. It addresses technological, managerial, political, economic and organizational aspects of the application of IT in relationship with Business Management. The journal will serve as a comprehensiveresource for policy makers, government officials, academicians, and practitioners. GJACST promotes and coordinates the developments in the IT based applications of business management and presents thestrategic roles of IT and management towards sustainable development.

# Global Journal of Advanced Computer Science and Technology

# Global Journal of Advanced Computer Science and Technology

**(Volume No. 11, Issue No. 2, May - August 2023)**

## Contents

# A Comprehensive Evaluation of Cryptographic Algorithms: DES, 3 DES, AES and Blowfish

**[1] Sangita Bhusare, [2] S B Thorat**

[1]Department of Computer Science and Information Technology, Yeshwant Mahavidyalaya, Nanded

[2]Institute of Technology & Management, Nanded

E-mail: [1]ymnbhusare@gmail.com, [2]suryakant_thorat@yahoo.com

## ABSTRACT

*Security is the most demanding aspects in the internet and network applications. Internet and networks applications are growing very fast, so the importance and the value of the exchanged data over the internet or other media types are increasing. Encryption is the process of scrambling a message so that only the planned recipient can read it. Encryption can present a means of securing information. Cryptography is an important part of modern world information security making the essential world a safer place. Cryptography is a process of making information incoherent to an unauthorized person. There are various cryptographic algorithms that can be used. In this procedure of choosing cryptographic algorithms, a study of strengths, weakness, cost and performance of each algorithm. In our paper, we have implemented and analyzed in detail cost and performance of popularly used cryptographic algorithms. This paper provides a comparison between four most common and used symmetric key algorithms: DES, 3DES,AES and Blowfish. A comparison has been made on the basis of these parameters: Time, Entropy and Avalanche effect, rounds, block size, key size, encryption/decryption time, CPU process time in the form of throughput and power consumption. Simulation program is implemented using Java programming.*

***Keywords - Encryption Algorithms, Cryptography, AES, DES, Blowfish, 3DES.***

## I. INTRODUCTION

Information Security is becoming much more vital in data storage and transmission. Hence the search for the best solution to offer the requisite protection against the data intruders‟ attacks along with providing these services in time is one of the most interesting subjects in the security related communities. Cryptography is the one of the main of computer security that converts information from its normal form into an unreadable form. The two main characteristics that identify and differentiate one encryption algorithm from another are its capability to secure the protected data against attacks and its speed and efficiency in doing so.

In today‟s world of gadget addiction, information storage, processing and recovery are computer based. The government, judiciary, small to big enterprises and almost every individual is using computer and internet based services. Online applications like shopping apps, banking apps, social networking apps provide services round the clock. But if an attacker gets a banking password, money is stolen at the same ease. An attacker may get social media login credentials and use it for mischievous activities. Hence,

securing information on computer, information sent via network, and information residing in applications is necessary. Cryptography is one such mechanism used in securing information and we will be analyzing cryptographic algorithms in our paper.

## II. ALGORITHMS IN OUR EXPERIMENT

### 2.1. DES

Data Encryption Standard (DES) is a symmetric key block cipher. The key length is 56 bits and block size is 64 bit length. It is vulnerable to key attack when a weak key is used. It began with a 64 bit key and then the NSA put a restriction to use of DES with a 56- bit key length, hence DES discards 8 bits of the 64 bit key and then uses the compressed 56 bit key derived from 64 bit key to encrypt data in block size of 64- bits .DES can operate in different modes - CBC, ECB, CFB and OFB, making it flexible. It is vulnerable to key attack when a weak key is used. In 1998 the supercomputer DES cracker, with the help of lakh‟s of distributed PCs on the Internet, cracked DES in 22h.

### 2.2 3DES

In cryptography, Triple DES is also called Triple Data Encryption Algorithm which is a block cipher. Three times to each block of data, Encryption – Decryption – Encryption using DES. The key length is 112 bits or 168 bits and block size is 64 bit length. Because of the increasing computational power available these days and weak of the original DES cipher, it was subject to brute force attacks and various cryptanalytic attacks; Triple DES was designed to provide a relatively simple method of increasing the key size of DES to protect against such attacks, without designing a completely new block cipher algorithm.

### AES

Advance Encryption Standard (AES) algorithm is a symmetric key block cipher. AES algorithm, supports any combination of data and key length of 128, 192, and 256 bits. AES allows a 128 bit data length that can be split into four basic operational blocks. These blocks are considered as array of bytes and organized as a matrix of the order of 4×4 which is also called as state and subject to rounds where various transformations are done. For full encryption, the number of rounds used is variable N = 10, 12, 14 for key length of 128,192 and 256 respectively. Each round of AES uses permutation and substitution network, and is suitable for both hardware and software implementation.

**Blowfish**

Blowfish is a symmetric key block cipher with key length variable from 32 to 448 bits and block size of 64 bits. Its structure is fiestal network. Blowfish is a symmetric block cipher that can be used as a informal replacement for DES or IDEA. It takes a variable- length key, from 32 bits to 448 bits, making it ideal for both domestic and commercial use. Blowfish is a fast, free alternative to existing encryption algorithms. From then it has been analyzed considerably, and it is slowly gaining popularity as a robust encryption algorithm. Blowfish is not patented, has free license and is freely available for all uses.

## III. IMPLEMENTATION

We have implemented and compared DES, 3DES, AES and blowfish with proposed approach. We have implemented the algorithms in java using „Net beans" IDE .We has used packages java security and java crypto. The packages java crypto and security provides security features like encryption, decryption, key generation, key management infrastructure, authentication and authorization features. We have used files of sizes 25KB, 50KB, 100KB consisting of text as input for encryption. The encrypted output of each file is saved as a file, which in turn is input for decryption. For sake of comparison we have used the same input files for all algorithms throughout the experiment. We have used a same system for all implementations and analysis work, so that memory and processor conditions remain same for all algorithms for comparison. Java crypto and security package contains the classes and interfaces that implement the Java security architecture. Using the libraries of these packages, we implemented and compared various cryptographic algorithms with proposed approach. The method of implementing algorithms using functions of java.security and javax.crypto package is as follows: Generate key using key generator class, create a cipher object with parameters algorithm name and mode, initialize the cipher created for encryption/decryption and perform encryption/decryption using doFinal() method.

## IV. EVALUATION PARAMETERS

Each of the encryption techniques has its own strong and weak points. In order to apply a suitable cryptography algorithm to an application, we should have knowledge regarding performance, strength and weakness of the algorithms. Therefore, these algorithms must be analyzed based on several features. In this paper, analysis is done with following metrics under which the cryptosystems can be compared are described below:

**Encryption time**

The time taken to convert plaintext to ciphertext is encryption time. Encryption time depends upon key size, plaintext block size and mode. In our experiment we have measured encryption time in milliseconds. Encryption time impacts performance of the system. Encryption time must be less making

**Avalanche effect**

In cryptography, a property called diffusion reflects cryptographic strength of an algorithm. If there is a small change in an input the output changes significantly. This is also called avalanche effect. We have measured Avalanche effect using hamming distance. Hamming distance in information theory is measure of dissimilarity. We find hamming distance as sum of bit by bit xor considering ascii value, as it becomes easy to implement programmatically. A high degree of diffusion i.e. high avalanche effect is desired. Avalanche effect reflects performance of cryptographic algorithm. Avalanche effect =( hamming distance ÷ file size) (1)

**Entropy**

Randomness is an important property in cryptographic processes because information should not be able to be guessed by an attacker. Entropy is measure of randomness in the information. It measures uncertainty in the information. In information security, we require security algorithms to yield high randomness in encrypted message, so that there is less or no dependency between key and cipher text. With high randomness, the relationship between key and cipher text becomes complex. This property is also called confusion. A high degree of confusion is desired to make it difficult to guess to an attacker. Entropy reflects performance of cryptographic algorithm. We calculate entropy using Shannon‟s formula.

## V. RESULTS AND DISCUSSIONS

Simulation results obtained after comparison for above parameters shows encryption times (in Milliseconds) of Encryption Algorithms with different input file Size. The result shows that the proposed algorithm required less time. Entropy values of Encryption Algorithms. The result shows that the proposed algorithm has good entropy. And Avalanche effect of Encryption Algorithms. The result shows that the proposed algorithm has good avalanche effect also.

## VI. CONCLUSION

The proposed algorithm uses mathematical concept of convolution sum and deconvolution with the combination of Z-Transform for encryption and decryption for providing strong security in message transmission by adding more complexity as compared to classical substitution and transposition cipher algorithms to increase Confusion and Diffusion in Cipher text. Proposed system will be an effective technique for the applications securing short messages while transferring it on internet.

## REFERENCES


[1] William Stallings, "Cryptography and Network Security: Principles & Practices", Fifth edition, Prentice Hall, ISBN- 13: 978- 0136097044, 2010.

[2] Bernard Menezes, "Network Security and Cryptography", Third edition, Cengage ISBN-13:9788131513491,2014.

[3] Dimitris G Manolakis, John G. Proakis, "Digital Signal Processing : Principles, Algorithms, and Applications", Fourth edition, PEARSON, ISBN: 9788131710005, 8131710009, 2007.

[4] Shamir, A.: How to share a secret. Commun. ACM 22(11), 612–613 (1979)

[5] Sharma, A., Singh, A.: Hybrid improved technique for data security and authentication for RDIF tags. In: IEEE International Conference on Signal Processing, Computing and Control (ISPCC2017)

[6] Shahara Banu, T., Shajeesh, K.U.: Secure reversible data hiding technique on textures using double encryption. In: International Conference on Innovations in Information, Embedded and Communication System (ICIIECS 2017)

[7] Tim Mather, Subra Kumaraswamy, and Shahed Latif "Cloud Security and Privacy", O"Reilly Media, Inc, pp 61- 71, 2009.

[8] Mohit Marwaha, Rajeev Bedi, "Applying Encryption Algorithm for Data Security and Privacy in Cloud Computing", IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 1, pp 367-370, 2013.

[9] V.U.K. Sastry, N. Ravi Shankar and S. Durga Bhavani, "A Generalized Playfair Cipher involving Intertwining, Interweaving and Iteration", International Journal of Network and Mobile Technologies, pp 45-53, 2010.

[10] Quist-Aphetsi Kester, "A Hybrid Cryptosystem Based on Vigenere Cipher and Columnar Transposition Cipher", International Journal of Advanced Technology & Engineering Research (IJATER), Volume 3, Issue 1, pp 141-147, 2013.

[11] Nagaraj, S., Raju, G., Srinadth, V.: Data encryption and authetication using public key approach. Procedia Comput. Sci. 48, 126–132 (2015)

[12] Hellman, M.: A cryptanalytic time-memory trade-off. IEEE Trans. Inf. Theory 26(4), 401–406 (1980)

[13] Lee, G.W., Hong, J.: Comparison of perfect table cryptanalytic tradeoff algorithms. Des. Codes Cryptogr. 80(3), 473–523 (2016)

[14] Oyetola Oluwadamilola, K., Okubanjo Ayodeji, A., Osifeko Martins, O.: An improved authentication system using hybrid of biometrics and cryptography. In: International Conference on Electro-Technology for National Development (NIGRCON) (2017)

[15] Harini, M., Pushpa Gowri, K., Pavithra, C., Pradhiba Selvarani, M.: A novel security mechanism using hybrid cryptography algorithm. (ICEICE2017) .

[16] Rachmawanto, E.H., Amin, R.S., Setiadi, D.R.I.M., Sari, C.A.: A performance analysis Stego Crypt algorithm based on LSB-AES 128 bit in various image size. In: International Seminar on Application for Technology of Information and Communication (iSemantic2017)

[17] D"souza, F.J., Panchal, D.: Advanced encryption standard (AES) security enhancement using hybrid approach. In: International Conference on Computing, Communication and Automation (ICCCA2017)

[18] Ebrahim, M.A., El-Maddah, I.A.M., Mohamed, H.K.: Hybrid model for cloud data security using steganography. @2017 IEEE

[19] Tian, Y., Gu, D., Gu, H., Ding, N.: Improved cryptanalytic of time-memory trade-off based on rainbow table. In: ICINS 2014, 2014 International Conference on Information and Network Security, Beijing, pp. 97–104 (2014)

[20] Hong, J., Moon, S.: A comparison of cryptanalytic tradeoff algorithms. J. Cryptol. 26(4), 559–637 (2013)

[21] Mohit Marwaha, Rajeev Bedi, "Applying Encryption Algorithm for Data Security and Privacy in Cloud Computing", IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 1, No 1, pp 367-370, 2013.

[22] Priyadarshini Patila, Prashant Narayankar, Narayan D G, Meena S Md "A Comprehensive Evaluation of Cryptographic Algorithms: DES, 3DES, AES, RSA and Blowfish", International Conference on Information Security & Privacy (ICISP2015), 11-12 December 2015, Nagpur, INDIA.

[23] C. Giolli, A. Scrivani, G. Rizzi, F. Borgioli, G. Bolelli, and L. Lusvarghi, "Failure mechanism for thermal fatigue of thermal barrier coating systems", Journal of Thermal Spray Technology,vol.18,pp.223–230,2009.

*[24] C. Zhou, Q. Zhang, and Y. Li, "Thermal shock behavior of nanostructured and microstructured thermal barrier coatings on a Fe-based alloy", Surface & Coatings Technology,vol.2170–75,2013.*
*[25] Sangita Bhusare, Dr.Suryakant Thorat," Symmetric Encryption Algorithm for Data Security and Privacy using Linear Convolution Sum Techniques", International Conference on Science, Technology ,Engineering and Management (ICSTEM), AE-ICSTEMBOSTON 16108- 001.*

# Project Assignment Problem Solved by Neuro- Fuzzy Inference System

**[1] Zoran Gacovski, [2] Ilija Gjorgjiev**
[1]European University-Skopje, bul. Kliment Ohridski, 68, Skopje, Macedonia
[2]Ministry of Information Society and Administration - Skopje
E-mail: [1]zoran.gacovski@eurm.edu.mk, [2]ilegorgiev@yahoo.com

## A B S T R A C T

*The project assignment problem is one of the most important decision issues for the managers in the large corporations. In this paper we propose a new neuro-fuzzy inference system that will help human decision-makers. We show that human-like decisions can be achieved with the proposed method. ANFIS (adaptive neuro-fuzzy inference system) is a method for fuzzy modeling that learns information about a data set, in order to compute the membership functions that best allow the associated fuzzy inference system to track the given input/output data. We have applied two methods for generating the rule base – starting from empty rule base (command line version) and starting with initial rule base (GUI version). For the training and testing sets – we have used student polls about job preference. Finally, we have realized simulation in Matlab, and solved different examples of decision-making for the project assignment problem.*

***Keywords - Neuro-Fuzzy Inference System, Project Assignment Problem, Adaptive Systems, Uncertain Decision-Making.***

## I. INTRODUCTION

The composition of project teams is one of the key roles of the project managers in the large corporations. It is an essential task to assign the right job to a right person. There are many factors that have impact on the assignment process. Sometimes – conflict criteria appear that make the decision process more complex. In many companies – the managers are doing these assignments manually – without taking into consideration the personnel expertise, their experience, type of the project, employees' preferences etc. Our work is a try to create an automated tool that will assist during the project assignment process, i.e. intelligent decision-making technique. The job assignment problem (JAP) is non- trivial and known to be a member of the class of NP- hard problems, and therefore in many cases it is even very difficult to find a "reasonable solution in a reasonable time". Also, this problem is omnipresent in and highly relevant to many industrial application domains like product manufacturing and workflow organization. This is because the JAP constitutes the core of most scheduling tasks, and the effectiveness and efficiency of whole companies and organizations is therefore often considerably affected by the way in which they solve this problem in its concrete appearance. The basic variant of the JAP studied here requires assigning jobs to executing staff such that the overall completion time is reduced, where there may be dependencies among the individual jobs and differences in the execution

abilities of the individual workers. Obviously, the JAP inherently allows parallelism and distributivity, simply because the jobs to be executed can be distributed over several people, and, because the workers can execute different jobs in parallel. There are several attempts to employ intelligent techniques for solving the job assignment problem – Jain et all. [7] used learning machines,

Kondadadi et all. [1] applied evolutionary computing, while Liang et all. [8] used neural nets for US Navy personnel assignment model. We propose here – an adaptive neuro-fuzzy inference system (ANFIS) which can learn to make human-like decisions and uses fuzzy membership functions for the soft constraints (input variables). For our study – we chose three widely used soft constraints. The goal is to find the most appropriate decisions resulting from fuzzy rules, for matches between projects and employees. Further goal is to extract a small number of fuzzy rules, which are still very effective without giving up robustness. Even though managers may face some difficulties in defining the real factors, the use of fuzzy membership functions is convenient, because they can easily express their decisions in terms of linguistic descriptions (low, medium, high) for the input constraints, as well as for the overall match between the projects and the employees. However, some managers may provide fairly diverse decisions, due to their personal experience, not to mention emotions, subjective factors, mistakes etc. Moreover, manager decisions are highly correlated and virtually non-deterministic. One manager can make 20% difference in decision even if the same data would be presented to him at a different time. Such indeterminate subjective component makes the optimization a very sophisticated task, very suitable for artificial intelligence solution. This paper is organized as follows: in Section 2 – the problem of project assignment is described; in Section 3 – the design of an adaptive neuro-fuzzy inference system (ANFIS) is given. In Section 4 – the simulation results are presented, and in Section 5 – concluding remarks are provided.

## II. PROJECT ASSIGNMENT PROBLEM

In this section we'll give short overview of the project (job) assignment problem. The JAP as it isconsidered within the frame of the work described here can be formally described as follows. Let J = $\{J_1,\ldots, J_n\}$ be a set of jobs and N = $\{N_1,\ldots,N_m\}$ be a set of workers, where each job can be executed by at least one of the workers (n, m Î N). The problem to be solved is to find an assignment of the jobs to the workers such that the overall time required for completing all jobs contained in J is minimal. Because this problem is NP-hard, usually it is reformulated such that it is just required to find an almost optimal solution in polynomial time.

Here, for solution of the project assignment problem - we mainly consider supervised learning methods. In large corporations – usually all projects and all employees are put in two separate databases. First

automation of the decision process can be done if every employee is joined with all possible projects – that satisfy his hard constraints (preferences = desires). So – we can form unique group for each employee, and we'll denote it with a group id. The numbers of projects in each group will be counted and normalized to fit the unit interval [0, 1]; this will be used later for decision refinement. This is important because the outputs (decisions made by managers) are highly correlated: there is typically one project offered to each employee. The data can be presented to a manager, who will make decisions (project offers), based on the following three widely used soft constraints (input variables):

- Duration of the project (in weeks);
- Type of the project ( its complexity);
- Pay-grade match (the earning that the employee will get from the project).

Note that decisions can be made sequentially and independently. This simulates the typical task a manager normally faces: offer a project to a given employee at the given time from a given list of possible projects, not considering other employees, projects that will be available in future etc. Of course, the employee does not have to accept the offered project and various other things may happen until the project starts.

For our adaptive neuro-fuzzy inference systems, we've performed training using students polls about project preferences (30 students were questioned). Decisions for projects to be offered have been individually evaluated for the best result in our model, and it was determined that even the misclassified predictions would make sense as real human-like decisions based on the given soft constraints. Due to the fact that we use fuzzy membership functions and that sometimes – two or more jobs could equally or nearly equally satisfy the given soft constraints, the highest value may not be easily available. In our simulation – we would prefer to come up with a small set (from 1 to 4 possible projects) offered to a given employee. Actually – in real life, it is also acceptable to offer 3 projects to employees to choose from.

## III. ADAPTIVE NEURO-FUZZY INFERENCE SYSTEM (ANFIS)

The basic idea behind these neuro-adaptive learning techniques is very simple. These techniques provide a method for the fuzzy modeling procedure to learn information about a data set, in order to compute the membership function parameters that best allow the associated fuzzy inference system to track the given input/output data. This learning method works similarly to that of neural networks. The adaptive neuro-fuzzy inference system uses a given input/output data set, and constructs a fuzzy inference system (FIS) whose membership function parameters are tuned (adjusted) using either a back-propagation algorithm alone, or in combination with a least squares type of method. This allows your

fuzzy systems to learn from the data they are modeling. The parameters associated with the membership functions will change through the learning process. The computation of these parameters (or their adjustment) is facilitated by a gradient vector, which provides a measure of how well the fuzzy inference system is modeling the input/output data for a given set of parameters. Once the gradient vector is obtained, any of several optimization routines could be applied in order to adjust the parameters so as to reduce some error measure (usually defined by the sum of the squared difference between actual and desired outputs).

Applying a neuro-fuzzy inference system for the project assignment problem is quite natural, because of the similarity to real life human decision-making. Some of the design issues were the following:

- **Number of input membership functions:** the fuzzy membership functions were set up based on knowledge of the managers. We determined that three membership functions (low, medium, high) for each of the three input variables model manager's decisions well.

- **Type of input membership functions:** based on the properties of the input variables – we considered triangular membership functions, but trapezoid, Gaussian and bell-shaped MFs were tested as well.

- **Type of the output membership function:** we used a single output, obtained using weighted average defuzzification. All output membership functions had the same type and were either constant or linear.
- The output variable is "employee-project likelihood" and the number of output membership functions ranged from 2 to 81.

- **The number of rules:** for a well defined fuzzy system – we need to define fuzzy output for every possible combination of input MFs values. If the training set doesn't include examples for given combination of values, but the testing set does, then in the testing phase we may still guess the output value using the aggregate distancefunction from all the learned fuzzy rules. This can be done before the actual testing phase begins resulting in a fully defined fuzzy system. In our case (3 input variables and 3 MFs) – there are 27=33 fuzzy rules, where the linguistic values were not negated, and they were connected with "and" relation. However, through rule extraction and combination of rules – we will drastically decrease the number of rules.

- **Performance function:** some of the widely used performance functions in neural networks are Sum of the squared error and Mean squared error. To verify training performance – we can also verify the correct classification rate.

- **Optimization methods:** back-propagation and hybrid (mixed least squared and back-propagation) methods were used as optimization methods.

- **The data was partitioned into 3 sets:** training, testing and cross-validation: the range of the training data set size was 50-90%. The cross- validation and the testing data sets each took half of the rest of the data (5-25%). The use of cross- validation is optional, but in our simulation is important, to avoid over-training.
- Number of training epochs – in most runs it was set up to 50 epochs.

Through an adaptive neuro-fuzzy inference system, the range of the membership functions are learned, fuzzy rules are created, and their weights are adjusted in order to better model the training data. The performance function values are calculated and classification is provided.

## IV. SIMULATION RESULTS

Our simulation was made in Matlab/ Fuzzy Toolbox. We used triangular, trapezoid and Gaussian as input membership functions, but triangular performed best in comparison with other types of MFs. Linear output membership functions performed better than constant ones.

For optimization method – the back-propagation and hybrid method performed similarly regarding the performance function and classification, but the hybrid method's running time was about 5 times as long as that of back-propagation's.

We have applied two methods for generating the rule base – starting from empty rule base (command line version) and starting with initial rule base (GUI version). The obtained results in both cases were very similar (comparison is given bellow).

Fig. 1 shows the ANFIS structure with its generated rules. The layered structure from left to right is the following:

Layer 1 – input neurons, three input neurons for the three input variables.

Layer 2 – input membership functions – three triangular membership functions for each input neuron.

Layer 3 - fuzzy rule left hand sides, each connected to 3 input membership functions.

Layer 4 - output membership functions (right hand sides) – the right hand side rules are in one to one relation with the left hand side rules.

Layer 5 – aggregated output, each output membership function gets aggregated along with the weight they carry.

Layer 6 – Output (decision).

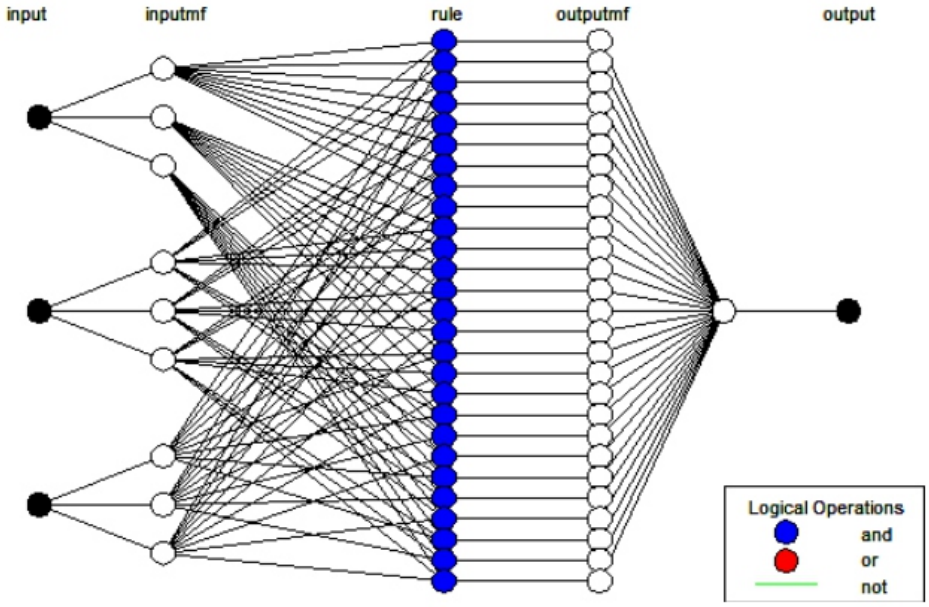


**Fig. 1. ANFIS structure with the generated rules.**

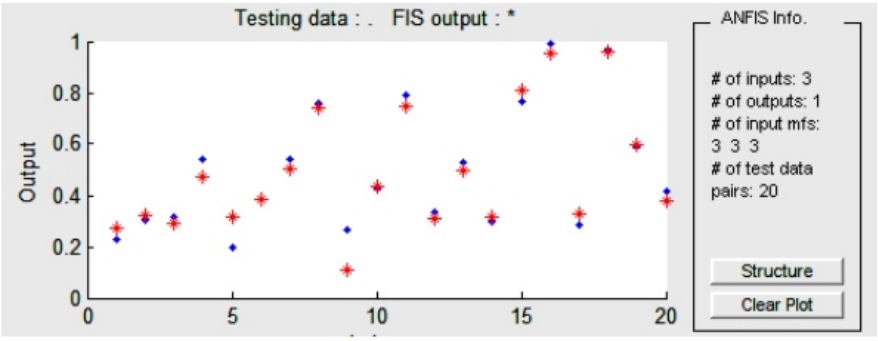


**Fig. 2. Testing data vs. actual FIS output.**

The training data are obtained from students and employees polls about their preference of certain projects. Table 1 shows the learned triangle membership function values after rounding. It is obvious that for all three constraints (input variables) – the leftmost membership functions maximum is close to 0. On Fig. 2 is shown the set of testing data vs. the real values of the FIS output.

| mf1 | mf2 | mf3 |
|---|---|---|
| dur_low= [-9.6 0 9.6] | comp_low= [-4 0 4] | pay_low= [-40 0 40] |
| dur_med= [2.4 12 21.6] | comp_med= [1 5 9] | pay_med= [10 50 90] |
| dur_hi= [14.4 24 33.6] | comp_hi= [6 10 14] | pay_hi= [60 100 140] |

**Table 1 – Membership functions after training**

| | Before training | After training |
|---|---|---|
| Largest difference | mf6= [0.0004 0.0001 0.0056 5e-05] | mf6= [-0.00052 -0.00029 -0.0046 -5.83e-05] |
| Smallest difference | mf1= [0.017 0.016 0.021 0.0045] | mf1= [0.01678 0.01062 0.0221 0.004419] |
| Average difference | mf19= [0.00012 .0007 0.001 1e-05] | mf19= [0 0 0 0] |

**Table 2 – Comparison of output values before and after ANFIS training.**

Further improvement of this research will be to extract a small number of rules, which can reliably predict projects to be offered based on fuzzy membership function values. Rule extraction can not further improve performance, but it can increase speed and efficiency for further training.

## V. CONCLUSION

In this paper – the project assignment problem is solved by an adaptive neuro-fuzzy inference system.

We showed that human-like decisions can be achieved with the proposed method. ANFIS (adaptive neuro-fuzzy inference system) is a method for fuzzy modeling that learns information about a data set, in order to compute the membership functions that best allow the associated fuzzy inference system to track the given input/output data. We have applied two methods for generating the rule base – starting from empty rule base (command line version) and starting with initial rule base (GUI version). For future research purposes – the number of fuzzy rules will be reduced, but it should held good performance as the original set of rules trained by the ANFIS for decision-making. For the training and testing sets – we have used employees' polls about job preference. Finally, we have realized simulation in Matlab, and solved different examples of decision-making for the project assignment problem.

## REFERENCES


*[1] R. Kondadadi, D. Dasgupta and S. Franklin, "An Evolutionary Approach for the Job Assignment Problem", Proc. of International Conference on Intelligent Systems, Louisville, Kentucky, 2000.*

*[2] R. Bayardo and R. Agrawal, "Mining the most interesting rules", in ACM KDD Conference, 1999.*

*[3] R. Kozma, M. Sakuma, Y. Yokoyama and M. Kitamura, "On the Accuracy of Mapping Backpropagation with Forgetting", Neuro-computing, Vol. 13, No. 2-4, pp. 295-311, 1996.*

*[4] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen and A. Verkamo, ,,Fast Discovery of Association Rules", in Advances of Knowledge Discovery and Data Mining, MIT press, pp. 307-328, 1996.*

*[5] C. M. Bishop, "Neural Networks for Pattern Recognition", Oxford University Press, 1995.*

*[6] F. Girosi, M. Jones and T. Poggio, "Regularization theory and neural networks architectures", Neural Computation, vol. 7, pp. 219-269, 1995.*

*[7] Jain, S., & Sharma, A. (1995). On aggre-gating teams of learning machines. Theoretical Computer Science A, 137(1), pp. 85-108.*

*[8] T. T. Liang and T. J. Thompson, "Application and Implementation – A large-scale personnel assignment model for the Navy", the Journal for Decisions Sciences Institute, vol. 18, No. 2, 1987.*

# Digital Games based Language Learning for Arabic Literacy Remedial

**[1] Noor Azli Mohamed Masrop, [2] Hafawati Ishak, [3] Ghazali Zainuddin, [4] Siti Rosilawati Ramlan, [5] Muhammad Sabri Sahrir**

[1,2]Faculty of Science and Information Technology, Kolej Universiti Islam Antarabangsa Selangor (KUIS)

[3]Faculty of Islamic Civilisation Studies, KUIS

[4]Faculty of Education, International Islamic University Malaysia, IIUM

[5]Faculty of Major Language Studies, Islamic Science University of Malaysia, USIM

E-mail: [1]noorazli@kuis.edu.my, [2]haffaishak@gmail.my, [3]ghazali@kuis.edu.my, [4]rosilawati@usim.edu.my, [5]muhdsabri@iium.edu.my

## ABSTRACT

*Digital game play is becoming increasingly prevalent. Its participant-players number in the millions and its revenues are in billions of dollars. As they grow in popularity, digital games are also growing in complexity, depth and sophistication. This paper presents reasons why games and game play matter to the future of education. Drawing upon these works, the potential for instruction in digital games is recognised. Previous works in the area were also analysed with respect to their theoretical findings. Hence, the authors in this study propose some existing Arabic language learning games intended for education of children. The analysis result shows that the majority of Arabic language learning games is limited to alphabet content. The overall presentation lacks of quality in term of graphics, animations, colors, and voice-over.*

***Keywords - Digital Game Play, Educational Games, Arabic Language, DGBLL, Gamification.***

## I. INTRODUCTION

Children spend many of their waking hours engaged in play[1]. Play contributes to cognitive development in a number of ways. It helps children to develop imaginary and memory which is essential for thinking about past, present and future [2]. They play games in classes, in their off-hours, even as part of their private contemplation. There is a growing body of literature that recognises the importance of games in children students' lives. Over the last several decades, new technologies have allowed digital media to create a multibillion- dollar entertainment industry commonly known as computer video and digital games.

In 2016, the U.S. computer and video game industry generated $30.4 billion in revenue, according to new data released by the Entertainment Software Association (ESA) and the NPD Group. This is total consumer spend figure includes revenues from all hardware, software, peripherals, and in-game purchases Separately, ESA highlighted that video game software revenue grew 6 percent from the 2015 level. In 2016, video game software revenue which includes physical packaged goods, mobile games,

downloadable content, subscriptions and other revenue streams, reached $24.5 billion up from 23.2 billion in 2015. Sales in the U.S. digital games market had grown from $43.3 billion in revenue in 2018, up to 18 percent from 2017[3]. Furthermore, retail sales of video games in U.S. reached a record US$7.4 billion in 2004.

In 2004, Malaysia's online gaming market was worth US$7 million in subscription revenue in 2019 up to 2 percent in 2018 [4]. This statistics doesn't yet include other type of games (non-online) that are pirated. The original game software market in Malaysia is almost non-existent due to the piracy of most game software in the market. As an example, Malaysian Ministry of Domestic Trade and Consumer Affairs' (MDTCA) has seized more than 110,000 copies of pirated computer and video games in a raid on one illegal duplication lab, which is capable of duplicating 2.4 million pirated discs per year. Looking at the seized volume of pirated software, we can presume that there are many digital games players locally[5].

The digital games industry is a multi-billion dollar industry, with games being developed for a variety of platforms, devices and emerging technologies Educators and trainers have looked at the multi- billion dollars computer and video game industry for inspiration. While the primary purpose of games is entertainment, the underlying design employs a variety of strategies and techniques intended to engage players in gameplay. The traditional school finds, at present, difficulties to maintain the concentration of the student and to foment its desire to learn. Faced with this reality, the use of educational games is one of the ways to motivate the student to participate actively in the activities[6]. This paper discusses the various aspects of digital games with a view to use this powerful medium to support children's language learning.

## II. THE GAMES

Games is a difficult concept to define. To date, computer games researchers are still debating the definition of games; the entry of educators into the fray complicates matters further[7] . However, in thispaper, the terms 'digital game', 'mobile game', 'computer games' and 'video games' are defined as digital applications that can be controlled by individuals or groups of players using a phone, tablet, computer or a video/console such as Playstation or Xbox machine.

Digital games are today an important part of most children's leisure and is increasingly becoming an important part of our culture as a whole. According to Entertainment Software Association (ESA), 28% of game players are children under 18 years old in USA[8]. Hence certain features of games can be adopted to make learning interesting and rewarding for children.

### A. Digital Games-Based Language Learning (DGBLL)

Recent interest in games and learning stems from debates about the role and practices of education in a new century, rather than from belief that young people find games motivating and fun, therefore, they should be exploited in educational contexts.

Definitions of game-based learning mostly emphasize that is a type of game play with defined learning outcomes [9]. According to Cheng and Whang[10] digital game-based learning (DGBLL) refers to the learning approach that incorporates educational content or learning designs into digital games. Research is mainly concerned with the development of related competencies and literacies during game play, the role of games in the formation of learning communities either while gaming, or related to game play . We need to create both a good game as well as a good learning experience. This can be achieved by either creating a game for the purpose from scratch or adapting an existing game to fit into an educational framework[11]. But according to Chen etc al [12] the literature still appears to lack a systematic framework for guiding the integration of language learning with contextual game-based environments. The positive impact and instructional advantages of DGBLL materials over printed course materials have been reported in relation to improvements in learners' listening skills [13][14], in vocabulary knowledge [15] and writing skills [16], [17] in communication, grammatical accuracy and writing skills [18] and in learners' general fluency, pronunciation and reading skills in the target language[13]. Along with language skills, positive results have also been reported in raising learners' intercultural awareness and intercultural communicative competence[19], [16] because serious games provided learners an invaluable opportunity to truly experience the target culture.

### B. Categories of Games

There have been many attempts to classify games. For example Ramsi [20]classify games using generic descriptors of the current popular types, which include (1) action, (2) adventure, (3) strategy, (4) simulation, (5) shooter, (6) sports, (7) role-playing, and (8) puzzle games. However these categories are not mutually exclusive. Many games fall into more than one category [21], such as being both adventure and combat games.

Based on the statistical results reported in 113 reviewed papers, Chang and Hwang [10] suggest that, for those new researchers who intend to adopt the mobile gaming approach, the three most frequently adopted game types (i.e., simulation game, role-play game and gamification) can be considered first.

## III. WHY DO PEOPLE PLAY GAMES?

Research on the motivations for games playing have been carried out across a number of disciplines. Being able to scrutinize the aspects of motivating play is important for designing games and gamified systems since the main purpose of gameful interaction is essentially to provide motivational affordances [22] .

Thomas Malone[23] identified three main ways in which games were able to motivate players: fantasy, challenge and curiosity. A survey by ESA found three main reasons for gameplay: i) connect them with their friends (55%), ii) it helps their family spend time together (46%), iii) provides mental stimulation or education. (75%) [24].

In the US, most of the children gameplay was supported by parents. Four main reasons parents play games with their children are fun for the entire family (88%), they're asked to (76%), , good opportunity to socialize with the child (76%) and good opportunity to monitor game content (59%) [25].

Presumably the fact that something does happen encourage players to proceed, and the quality of what happens in terms of user engagement is the factor that keeps them playing [26]. It was suggested that the degree of difficulty of gameplay is important for children to enjoy playing and the game must be neither too difficult nor too easy [27].

## IV. GAMES AND CHILDREN

Game approach in education or playing while learning is in the Malaysian preschool curriculum [28]. Various researches on children gameplay have been carried out. Systematic literature review of research studies on game-based learning and gamification conducted in Asian K–12 schools, that conducted by So and Seo [29] have shown that the reviewed articles advocated the positive efficacy of games on learning outcomes Vasalou et al [29] apply a social constructivist lens to DGBL for children whostruggle with literacy. The findings show that children spontaneously engage in 'game talk' regarding game performance, content, actions and experiences. While this game talk facilitates a strong sense of social engagement and playfulness, it also caters to a variety of new opportunities for learning by sparking tutor and student-initiated interventions. Ismail [30] carried out the research about document analysis that focus on game, emotional, cognitive and psychomotor element based on standard curriculum blueprint for early childhood education. The result shows that the developmental elements applied in the game were suitable with the preschool students age and it can achieve the objectives that outlined by the National Preschool Curriculum Standard.

## V. ARABIC LANGUAGE LEARNING

Although Arabic has been studied as subjects in primary and secondary schools, a large number of students in Malaysia still can not speak Arabic. Awang et al, [31] states that students do not have the confidence to use Arabic in and out of the classroom. According to Ghani et al, [32] the weakness of the Arabic language has resulted in the impairment of the quality of teaching and learning of Arabic language in schools in Malaysia. Researchers have identified that among the problems faced by students in Arabic learning or other second language is related to the vocabulary[32], [33].

### A. Arabic vocabularies learning

Vocabulary is an important aspect in language because it is the first step in learning any language. Without vocabulary, learning language skills such as listening, speaking, reading and writing can not be realized successfully[33]. In other words, vocabulary is an essential part of language and it is the first step in learning any language. Ashinida [34] and Ghazali et.al [35] verified that poor knowledge of Arabic vocabulary limits the ability of students to communicate, write, read and listen to materials in Arabic.[36][37].

Therefore, Noor et al., [38] states that in order to encourage students to learn Arabic, teachers need to introduce interesting learning strategies such as independent and active learning through the use of electronic materials. Thus, the focus is on learning Arabic vocabulary only as a first step in learning Arabic language throughdigital games that provides more convenient and attractive learning process.

### B. Digital Games-Based Arabic Language Learning

A huge number of studies have been devoted to various fields. Digital games based learning (DGBL) in the field of language education have been widely used due to their positive outcomes in learning and learner participation. Usability, motivation, flow state, affective engagement, and learning were determined, which revealed that serious games were frequently used with a high level of engagement. However, there are not many relevant DGBL studies in the context of Arabic language learning. There are few games for learning Arabic dedicate to children. Generally they are simplistic and tend to revolve around the same trivial idea in language learning [20].

Sahrir, MS, & Alias, NA[39] reported a positive perception of university students in learning Arabic online. To ensure the effectiveness of using digital games in education, there are three main components that should be considered: i) Pedagogy ii) DGBL elements iii) ARCS model. There are some existing online games that teach Arabic language to children like ALADDIN, Salaam Arabic, and Araboh.com. Although they showed success for children, they do not address the Learning Disabilities (LDs) nor the intellectual problems [40].

### C. Analysis of some existing games

The authors have reviewed some existing Arabic Language Learning games intended for education of children and chose the followinggames: Arabic Games: Word and Vocabulary ,Arabic Alphabet Kids ,Secil Bahasa Arab, and Bee Learning Arabic. Figure 1 describes their major characteristics.

| Game name | Description |
|---|---|
| Arabic Games(AGW) | The players only have multiple choice question. Player can learn and play with 40 level with different task)[42] |
| Arabic Alphabet Kids (AAK) | The players provide teaching Arabic alphabet but also to read and write basic words using the names and sounds of animal. [43] |
| Secil Bahasa Arab (SBA) | -This game provides : four modul to learn and games to practise the modull, voice and shape to learn. [44] |
| Bee Learning Arabic (BLA) | - This game gives the player to know and learn alphabet and provide guide for players [41] |

**Figure 1: Descriptions of Arabic Educational Games**

| Game name | Content[(i)] | Organization of the contents[(ii)] | Interactivity[(iii)] | Graphics[(iv)] | Assessment[(v)] |
|---|---|---|---|---|---|
| AGW | 2 | 2 | 2 | 2 | 1 |
| AAK | 1 | 1 | 1 | 1 | 2 |
| SBA | 1,2,3 | 2 | 1 | 1 | 1,2,3 |
| BLA | 1 | 1 | 2 | 3 | 1 |

**Table 2: Content Analysis of Arabic Educational Games**

**Notes :**

(i) Content : (1) Alphabet, (2)Words, (3)Numbers, (4)Pronunciation, (5)sentence, (6)Grammar (ii) Organization of the contents: (1) Randomly, (2) Organized (iii) Interactivity: (1) Sounds guide the player, (2) Boring (iv) Graphic (1) Colorful (2) Moderate (3) Poor (v) Assessment: (1) Multiple choice (2)Puzzle (3) Guessing

## VI. CONCLUSION

This paper discusses the digital games and their potential use in supporting learning, Games can provide powerful learning environments, especially for children. We also Analyze some existing games. The analysis result shows that the majority of Arabic Language learning games is limited to alphabet content. The overall presentation lacks quality in term of graphics, animations, colors, and voice-over. Moreover, none of Arabic games shows a systematic design process. They do not have the ability to engage the learner in the game.

## ACKNOWLEDGEMENT

The authors wish to thank to KUIS and Selangor State Government which has support this research under GPPSTI (Geran Penyelidikan dan Pembangunan Sains Teknologi dan Inovasi). Grant Scheme.

## REFERENCES

*[1] A. S. Lillard, "The development of play," Handb. Child Psychol. Dev. Sci., pp. 1–44, 2015.*

*[2] S. Ahmad, A. H. Ch, A. Batool, K. Sittar, and M. Malik, "Play and Cognitive Development: Formal Operational Perspective of Piaget's Theory.," J. Educ. Pract., vol. 7, no. 28, pp. 72–79, 2016.*

*[3] The ESA, "U.S. Video Game Sales Reach Record-Breaking $43.4 Billion in 2018," The Entertainment Software Association. [Online]. Available: http://www.theesa.com/article/u-s-video-game-sales-reach-record-breaking-43-4-billion-2018/. [Accessed: 11-Apr- 2019].*

*[4] Statista, "Online Games - Malaysia | Statista Market Forecast." [Online]. Available: https://www.statista.com/outlook/212/122/online- games/malaysia. [Accessed: 11-Apr-2019].*

*[5] Entertainment Software Association (ESA), "Malaysian Authorities Shut Down Pirate Game Lab," The Entertainment Software Association. [Online]. Available: http://www.theesa.com/article/malaysian-authorities-shut-pirate-game-lab/. [Accessed: 21-Apr-2019].*

*[6] C. A. C. Bürger and T. S. Ghisleni, "Education and games: educomunicative analysis on the implementation of games in teaching environments," Res. Soc. Dev., vol. 8, no. 4, p. 4684900, 2019.*

*[7] M. M. Noor Azli, "Pembelajaran Berasaskan Permainan Digital (Pbpd) Bagi Pendidikan Literasi Jawi Berasaskan Model Motivasi," PhD, Universiti Kebangsaan Malaysia, 2014.*

*[8] The ESA, "U.S. average age of video gamers 2018 | Statistic," Statista. [Online]. Available: https://www.statista.com/statistics/189582/age-of-us-video-game-players-since-2010/. [Accessed: 11-Apr-2019].*

*[9] J. L. Plass, B. D. Homer, and C. K. Kinzer, "Foundations of game-based learning," Educ. Psychol., vol. 50, no. 4, pp. 258–283, 2015.*

*[10] C.-Y. Chang and G.-J. Hwang, "Trends in digital game-based learning in the mobile era: a systematic review of journal publications from 2007 to 2016," Int. J. Mob. Learn. Organ., vol. 13, no. 1, pp. 68–90, 2019.*

*[11] S. Andreassen, "Zombies in the classroom. Video games for engagement in a new century of education," Master's Thesis, The University of Bergen, 2015.*

*[12] Z.-H. Chen, H. H.-J. Chen, and W.-J. Dai, "Using narrative- based contextual games to enhance language learning: A case study," J. Educ. Technol. Soc., vol. 21, no. 3, pp. 186–198, 2018.*

*[13] S. Bernert-Rehaber and G. Schlemminger, "Immersive 3D- Technologien optimieren das Fremdsprachenlernen:„EVEIL- 3D–Lernen in virtuellen Welten "," Babylonia, vol. 3, pp. 44– 49, 2013.*

*[14] M. Roy and G. Schlemminger, "Immersion und interaktion in virtuellen realitäten: Der faktor präsenz zur optimierung des geleiteten sprachenlernens," Z. Für Interkulturellen Fremdsprachenunterricht, vol. 19, no. 2, 2014.*

*[15] O. Kocaman and G. K. Cumaoglu, "The effect of educational software (denis) and games on vocabulary learning strategies and achievement," Egitim Ve Bilim, vol. 39, no. 176, 2014.*

*[16] R. M. Levy and M. G. O'Brien, "A virtual world for teaching German," Loading, vol. 1, no. 1, pp. 1–17, 2006.*

*[17] D. O. Neville, B. E. Shelton, and B. McInnis, "Cybertext redux: Using digital game-based learning to teach L2 vocabulary, reading, and culture," Comput. Assist. Lang. Learn., vol. 22, no. 5, pp. 409–424, 2009.*

*[18] A. Berns, M. Palomo-Duarte, J. M. Dodero, and C. Valero- Franco, "Using a 3D online game to assess students' foreign language acquisition and communicative competence," in European Conference on Technology Enhanced Learning, 2013, pp. 19–31.*

*[19] V. Guillén-Nieto and M. Aleson-Carbonell, "Serious games and learning effectiveness: The case of It's a Deal!," Comput. Educ., vol. 58, no. 1, pp. 435–448, Jan. 2012.*

*[20] F. M. A. Ramsi, "A Game-Based Learning Model," Master, United Arab Emirates University, 2015.*

*[21] M. Prensky, Digital game-based learning. New York: McGrawHill, 2001.*

*[22] J. Koivisto and J. Hamari, "Demographic differences in perceived benefits from gamification," Comput. Hum. Behav., vol. 35, no. 2014, pp. 179–188.*

*[23] T. W. Malone, "What makes things fun to learn? A study of instrinsically motivating computer games.," 1980.*

*[24] Entertainment Software Association, "2018 sales, demographic, and usage data essential facts about the computer and video game industry," 2019.*

*[25] "Essential-Facts-2016.pdf."* .
*[26] J. Kirriemuir and A. Mcfarlane, Literature Review in Games and Learning. 2004.*
*[27] A. McFarlane, A. Sparrowhawk, and Y. Heald, Report on the educational use of games. TEEM (Teachers evaluating educational multimedia), Cambridge, 2002.*
*[28] Pusat Perkembangan Kurikulum, "Kurikulum prasekolah kebangsaan." Kementerian Pendidikan Malaysia, 2001.*
*[29] A. Vasalou, R. Khaled, W. Holmes, and D. Gooch, "Digital games-based learning for children with dyslexia: A social constructivist perspective on engagement and learning during group game-play," Comput. Educ., vol. 114, pp. 175–192, 2017.*
*[30] I. M. Ismail, S. Hashim, S. K. Anis, A. Ismail, and M. E. Ismail, "Implementation of a development in cognitive, psychomotor and socio emotional elements through games to achieve national preschool curriculum standards," in 2017 IEEE 9th International Conference on Engineering Education (ICEED), Kanazawa, 2017, pp. 143–148.*
*[31] N. A. Awang, M. H. Mohamed, and R. Sulaiman, "Enhancing Arabic Speaking Skills among Malay Students through Group Work Activities," Int. J. Humanit. Soc. Sci., vol. 3, no. 21, pp. 212–219, 2013.*
*[32] M. T. A. Ghani et al., "A questionnaire-based approach on technology acceptance model for mobile digital game-based learning," J. Glob. Bus. Soc. Entrep. GBSE, vol. 5, no. 14, 2019.*
*[33] E. K. Alobaydi, R. Y. Alkhayat, M. R. M. Arshad, and E. R. Ahmed, "Context-aware ubiquitous Arabic vocabularies learning system (U-Arabic): A framework design and implementation," in 2017 7th IEEE International Conference on Control System, Computing and Engineering (ICCSCE), 2017, pp. 23–28.*
*[34] A. Aladdin, "Analisis penggunaan strategi komunikasi dalam komunikasi lisan Bahasa Arab," GEMA Online® J. Lang. Stud., vol. 12, no. 2, 2012.*
*[35] G. Yusri, N. M. Rahimi, P. M. Shah, and W. H. Wah, "Cognitive and metacognitive learning strategies among Arabic language students," Interact. Learn. Environ., vol. 21, no. 3, pp. 290–300, 2013.*
*[36] A. Aladdin, "Analisis penggunaan strategi komunikasi dalam komunikasi lisan Bahasa Arab," GEMA Online® J. Lang. Stud., vol. 12, no. 2, 2012.*
*[37] G. Yusri, N. M. Rahimi, P. M. Shah, and W. H. Wah, "Cognitive and metacognitive learning strategies among Arabic language students," Interact. Learn. Environ., vol. 21, no. 3, pp. 290–300, Jun. 2013.*
*[38] Z. A. M. Noor, N. M. R. N. Yusoff, I. M. M. Yasim, and M. Y. Kamarudin, "Foreign Language Vocabulary Learning Strategies in Malaysia," Creat. Educ., vol. 7, no. 03, p. 428, 2016.*
*[39] M. S. Sahrir and G. Yusri, "Online vocabulary games for teaching and learning Arabic," GEMA Online® J. Lang. Stud., vol. 12, no. 3, 2012.*
*[40] J. Salah, S. Abdennadher, C. Sabty, and Y. Abdelrahman, "Super alpha: Arabic alphabet learning serious game for children with learning disabilities," in Joint International Conference on Serious Games, Springer, 2016, pp. 104–115.*
*[41] X-Gaf Studio, Bee Learning Arabic Kids - Apl di Google Play. 2018.*

# A New Method for Querying XML Data

**Alhadi A. Klaib**
Faculty of Information Technology, University of Elmergib
E-mail: Alhadi.klaib@elmergib.edu.ly

## A B S T R A C T

*Extensible Mark-up Language (XML) is currently an important standard method for formatting and exchanging data over the Internet. Updating and retrieving a massive amount of XML data is an interesting and active research area. In addition, indexing XML data is an important task to improve the efficiency of XML queries. Therefore, this paper surveys the methods that used for indexing XML data. Subsequently, a new approach is proposed for indexing and querying XML data efficiently. In addition, this approach used the idea of dividing the nodes of an XML document into groups and labelling them accordingly. Two existing labelling schemes were selected carefully for labelling these groups and their nodes in order to improve the efficiency of the XML query.*

***Keywords - XML Data; Indexing XML Data; Labelling Scheme; Query XML Data.***

## I. INTRODUCTION

Extensible Mark-up Language (XML) has become very significant technology for transferring data through the world of the Internet [1, 2]. Thus, a large amount of research on XML databases and XML technologies such as data retrieval and data update has being carried out [2]. XML labelling scheme is an important technique used to handle XML data efficiently and robustly. Labelling XML data is performed by assigning labels to all nodes in that XML document. Every node is provided with a unique label that can be used to build the relationship among nodes in that XML tree [3, 4]. Initially, the main focus of the XML research was about handling static documents in terms of data retrieval and navigation. [5-8]. Currently, most XML documents are dynamic and it is essential that they are handled efficiently since most database applications nowadays include XML processing. As a result, labelling XML data has become an important task to improve query processing [9]. The key challenges are with dynamic XML data since static XML data has been efficiently processed by proposing a number of successful schemes such as the Dewey scheme and containment scheme [10, 11]. The case with dynamic XML data is different as the XML databases still struggle to manage large numbers of relabeling cases. Many labelling schemes have been proposed for dynamic XML data [12-16]. Any efficient labelling scheme should provide effective query performance, labelling XML data efficiently, reducing the required relabeling cases, and determining the relationships. Therefore, this paper addressed this issue and came up with a proposal as a new approach for XML labelling scheme.

## II. OVERVIEW OF INDEXING XML DATA METHODS

A considerable amount of research has been conducted on querying and storing XML data since the importance of XML data `management has been increasing dramatically. Thus, there is a demand for and focus on producing an efficient labelling scheme. Therefore, enhancing XML efficiency by developing a robust XML scheme is mostly achieved by cutting the cost of data searching. The main purpose of the labelling scheme is to encode the information about the XML tree and structure in a very compacted label. The compactness of the labels and the performance speed of the scheme are the key metrics for any labelling scheme. A significant amount of research has been carried out to produce an efficient XML labelling scheme [4, 17-20].

XML sets the standard rules which are readable by both machine and users. XML has been recognised as a standard means of exchanging data on the WWW. Problems were expected to arise in the use of XML. Eventually these problems appeared and are mostly concerned the query performance [13, 21-23]; [1]. The query definition in this context is to allow users to access the XML data. Consequently, many XML query languages have been proposed. Some of these languages have been developed by the W3C group such as $X_{Query}$ and $X_{Path}$ [24]. These languages have features that assist in overcoming the disadvantages of XML. However, XML query performance still needs to be improved in order to always cope with the growth of the web. Therefore, developing an efficient indexing system has been carried out for many years [14, 25]. The existence of labelling schemes is an important issue for XML indexing since it allows queries to avoid going through the whole XML document as Murata et al. said [26]. They also stated that it is a significant step to assign a label for each node in the XML document in order to be able to determine the relationship among these nodes. The function of the nodes is to play the role of unique IDs for the components on the XML documents. Thus, labelling is required in order to execute structuralqueries which can only be performed by using the index [12]. As a result, the query process does not require access to the whole XML document, so that makes the query process efficient and quicker [27]. The size of the index is mainly based on the size of the labels; thus, many researches have concentrated on proposing labelling schemes to achieve this goal. Some other researches have proposed schemes that were developed to assist in path indexing and numbering schemes to ease the XML query. Some other schemes have focused on the ease of the function. For instance, a scheme was proposed by Bruno [28] to ease the compact method of the results of a query path and then the gathering of these paths produces the final combination for a twig query. This scheme was enhanced by Lu [29] to support queries efficiently. The improvement was on handling twig queries. A complete servey for labelling XML data techniques was accomplished and published [30].

## III. RESEARCH APPROACH

Theoretically, the idea of the proposed labelling scheme is to divide the whole data tree into small groups (clusters). This mechanism makes each cluster itself a sub-data tree. The advantage of this mechanism is to reduce the number of required re-labelling cases to the lowest possible level, and as a result improve the efficiency of the query performance processing. Subsequently, two XML labelling schemes were used to label the nodes and their clusters of a data tree. These two schemes are the Dewey labelling scheme which was used to label the clusters, and the LLS labelling scheme which was used to label the nodes of the data tree. This idea is used in order to ease the determination of the child–parent and sibling relationships as the child–parent relationship is available in each XML document. Furthermore, these relationships also ease the process of inserting new nodes. The parent–child clustering-based technique helps in dealing with a small tree rather than the entire XML tree [31]. It was also found that the parent–child clustering technique supports the labelling process, and is more efficient than the simple tree in terms of the accuracy and spent time for query processing [32]. One of the features of a clustering-based technique is that it uses two labels for every node as this idea eases the procedure of labelling nodes that share the same cluster; whereas the label of the cluster is used to link that cluster with the entire tree. This feature assists in determining the relationships among nodes that form different clusters [15]. It was stated that a scheme is developed to provide fast identification of relationships, as this feature helps in the optimizing of query processing [33]. Therefore, these advantages support the proposed scheme, which consequently helps in enhancing the query processing and other targeted features of the proposed scheme such as improving the process of labelling XML documents. Figure 1 illustrates an example for the proposed scheme for a XML data document.

## IV. THE MECHANISM AND DATA MODEL OF THE PROPOSED SCHEME

The mechanism of the proposed method is carried out in two steps as follows: first of all, labelling XML Document: the implementation of this task is divided into two jobs: 1) creating and labelling clusters. All nodes are grouped into clusters. The mechanism for achieving this task is as follows:

A- Each node and its child nodes are gathered to build a cluster.

B-Only the main root of the document is considered as a cluster itself and without child nodes. This node is labelled number (1).

C-Each cluster is considered a sub-tree.

D- A cluster can have only one or two levels of nodes.

E-Each cluster has at least one node.

F- Each cluster may have a root and child node(s) that is connected with this root. Thus, a cluster must contain at least one node.

G- All clusters, including the main root (not each node), are labelled by the Dewey labelling scheme.

1) Secondly, labelling Nodes: all nodes are labelled according to the following mechanism:

a) Each cluster is treated as a sub-tree and its nodes are labelled separately from other clusters.

b) The LLS labelling scheme is used to label all nodes.

c) If the root of a cluster is a child node of another cluster, which means that this root has already been labelled, then the label of this root will be kept the same.

Figure number 2,3,4,5, and 6 show different scenarios of inserting new nodes into the proposed scheme and the any changes occur accordingly.

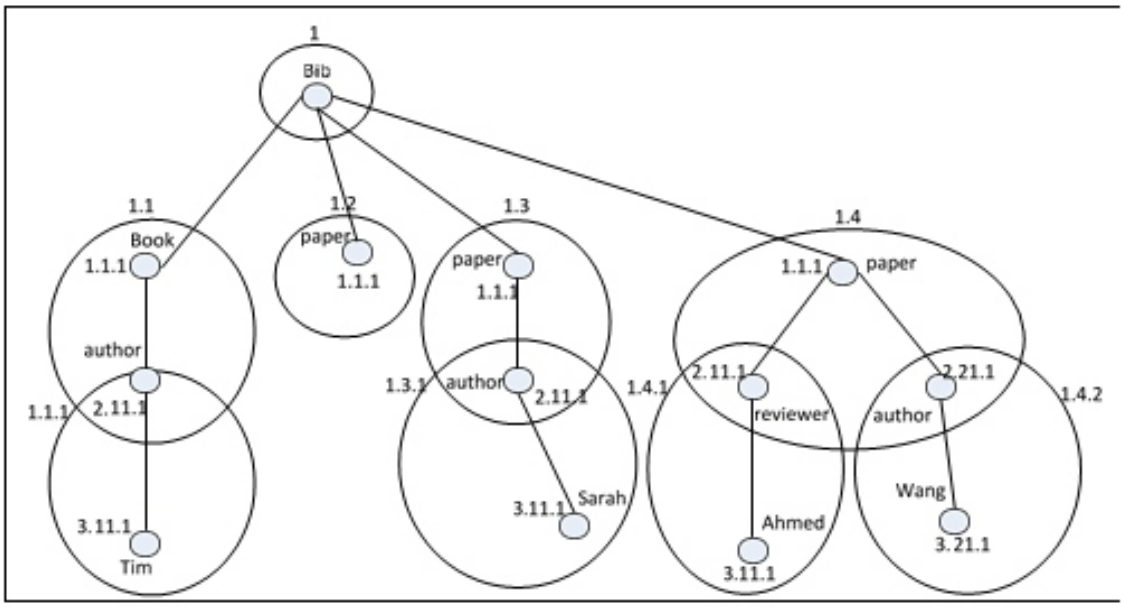


**Fig 1: an example for the proposed labelling scheme**

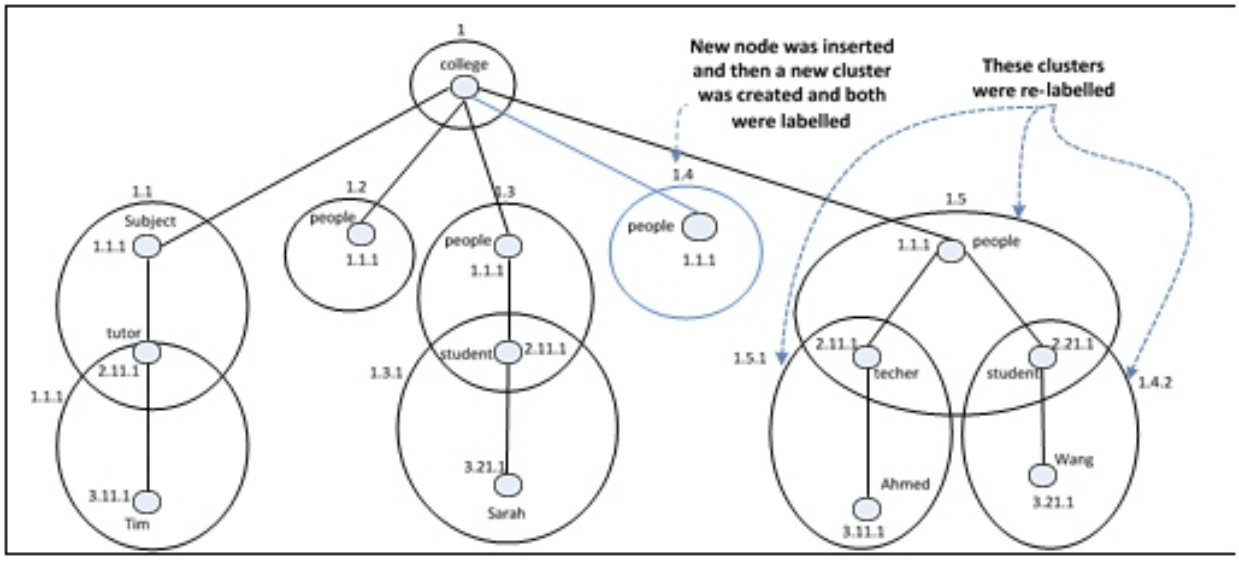


**Fig 2: Inserting a new node in the second level**

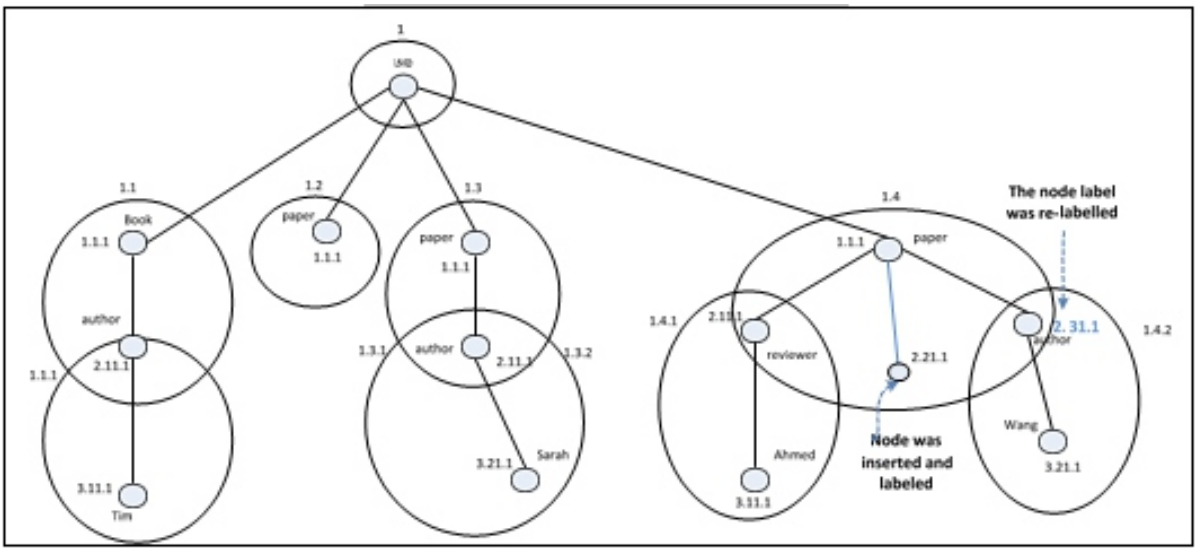


**Fig 3: Inserting a new node inside a cluster**

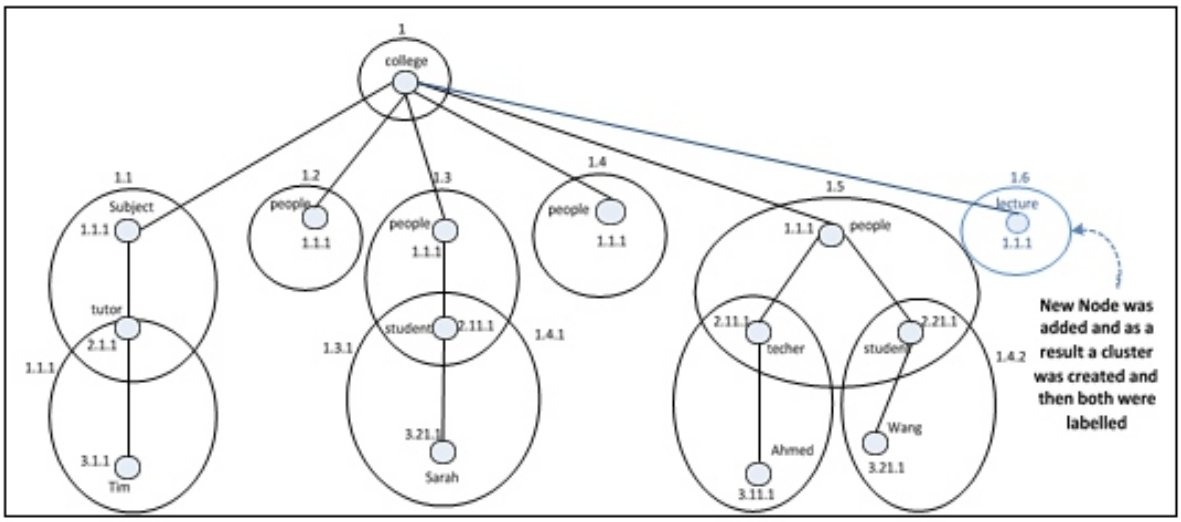


**Fig 4: Inserting a new node in the second level and rightmost side**

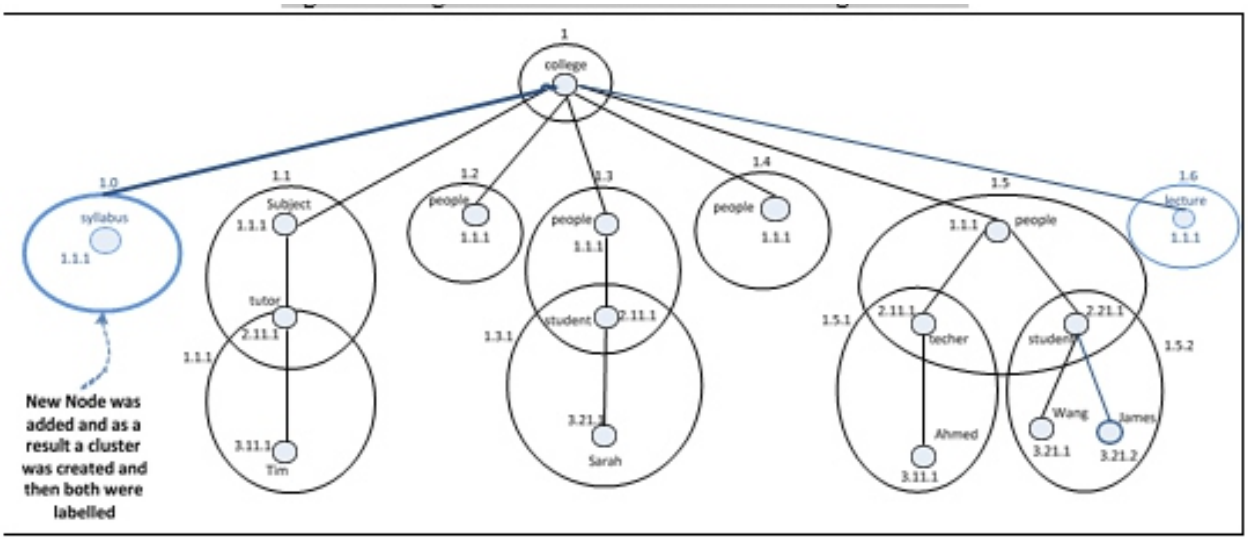


**Fig 5: Inserting a new node in the second level and leftmost side**

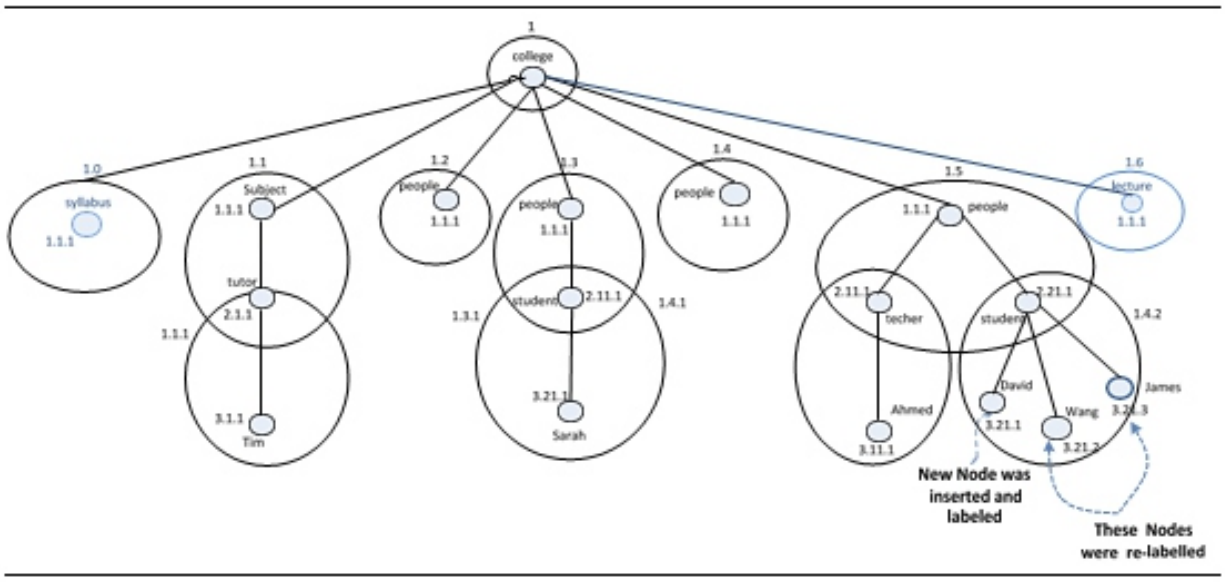


**Fig 6: Inserting a new node in the leftmost side within a cluster**

## V. RESULTS

This research came up with the idea of developing a state-of-the-art approach. The idea of this approach is to propose a hybrid labelling scheme using two existing labelling schemes.

Regarding the update cost of the proposed labelling scheme, the update processes that take place in the XML documents affect the databases and indexes. Thus, the proposed scheme aims to reduce the update cost to the lowest possible level. Reducing the number of required relabeling processes is the main technique to reduce the update cost. Using the clusters technique helps in reducing these relabeling processes. The following number of scenarios for the update cost for the proposed scheme against the LLS scheme and the Dewey labelling scheme show how the update cost was improved.

## VI. SCENARIOS FOR COMPARING THE PROPOSED SCHEME WITH OTHER SCHEMES

Figure 7 demonstrates an example of inserting a new node. This node is inserted into a cluster that has only one node and the root is a child of the main root. The figure also shows the required relabelling nodes for the proposed scheme and the LLS and Dewey labelling schemes. In this scenario, the proposed scheme can provide better performance than the others as there is no need for any relabelling in the proposed scheme, whereas relabelling operations are required in the other two schemes. The result of this scenario is a common case with the proposed scheme.

The worst scenario for the proposed scheme occurs when a new node is inserted between two nodes in the second level. Thus, a new cluster needs to be created and then some of the same level clusters need to

be relabelled. In this case the relabelling operations in the proposed scheme will be the same as the LLS scheme, and better than the Dewey scheme. However, this scenario is an uncommon case. Figure 8 shows an example of this scenario.s

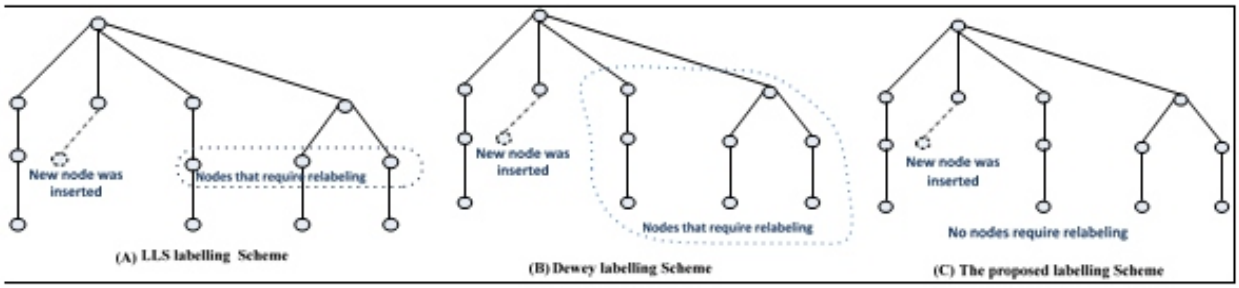


**Figure 7 : A scenario for comparing the proposed labelling schemes with the LLS and Dewey schemes**

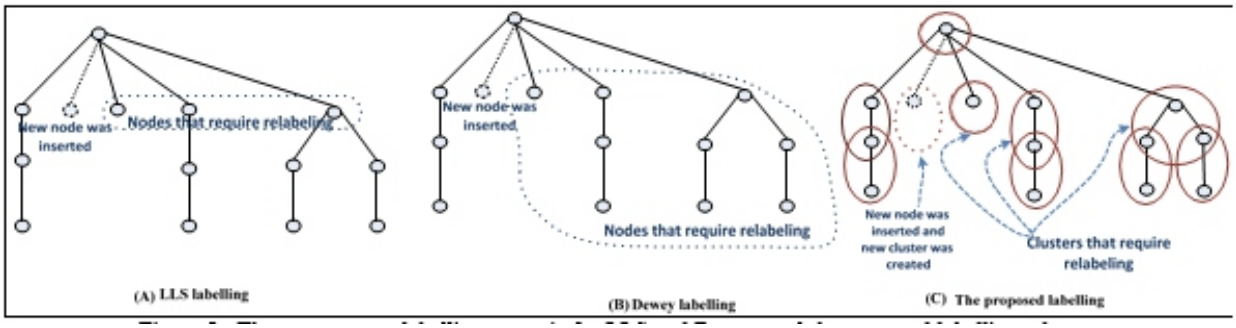


**Figure 8 : The worst case relabelling scenario for LLS and Dewey, and the proposed labelling schemes**

## VII. EXISTING APPROACHES AND OUR APPROACH, A COMPARISON AND JUSTIFICATION

Number of approaches have been proposed and developed as solutions for indexing XML data. Each approach has advantages and limitations and at the same time has a specific technique. This section discusses and compares the most common techniques used in these labelling schemes to resolve the targeted issues, in order to ascertain the limitations of these techniques and justify the improvement that the proposed scheme might provide.

First of all, let's start with the node schemes. The main idea of this technique is to save numbers that represent the positions of the nodes in the tree. Thus, accessing any node can be achieved by using these numbers. This technique is suitable for both single and twig queries. Secondly, the technique of a Graph indexing scheme is based on producing a path summary for the XML data. The path summary accelerates the process of query evaluation in order to improve query performance. There are a number of Graph schemes that have been proposed. Different techniques were employed in these schemes.

Thirdly, regarding the techniques of Sequence indexing schemes, they convert both XML documents and queries into structure sequences. Sequence indexes put the values and the structures of XML data together into an integrated index structure. This structure is used to evaluate both path and twig queries efficiently, answering a query, making a string sequence that matches th sequence of the data with the query. The problem with most sequence techniques is that they return non-precise and non-complete answers.

Concerning the technique of dividing the nodes into groups and introducing sub-trees, this technique has already been used because it has many advantages such as it decreases the re-labelling cases and therefore improves the insertion process. Improve the determination of the relationships [31]. In addition, the clustering technique has two labels for every node, and this feature improves the process of labelling nodes that from the same cluster; and the label of the cluster is used to connect that cluster with the whole tree. This feature helps in determining the relationships between the nodes that form different clusters. It is also simplify many scenarios of inserting new nodes [15]. However, despite these features of the clustering technique, it has the disadvantage that implementing two labels is a complicated process [34]. It was also observed that the parent–child clustering technique improves the labelling process, and better than the simple tree with regard to the accuracy and spent time for query processing [32]

This research, as many others in this field, tries to introduce a new technique for labelling XML data that can occupy the minimum storage space and at the same time keep the query performance efficiency at the best level. Therefore, this research proposes a labelling scheme that takes advantage of some of the existing schemes, using the clustering-based technique in order to introduce a technique that achieves a better solution than others. Consequently, this technique is based on the following main points:

- The Dewey labelling scheme was selected for labelling the clusters due to the advantages of this scheme such as the ability to manage the clusters, the ease of updating, and because it performs well.
- The LLS labelling scheme was selected for labelling the nodes of the clusters. The advantages of this scheme are that it uses numerical data and the size of the labels is steady.
- The clustering-based technique was used due to the great advantages of this technique. It reduces the required relabelling cases and as a consequence improves the inserting new nodes processes, with easy determination of the relationships.

All three of these ideas were intended to present the state of the art scheme that would achieve the aim of this research.

The model of the LLS scheme is based on the levels of the tree. This feature makes the query process of this scheme very fast. Additionally, the model of this scheme has a fixed size of labels regardless of the size of the tree. The Dewey model is a very common one and has been studied and adopted in many later schemes. It is a very robust scheme that supports the structure of the index. The decision to choose the LLS and Dewey schemes was considered carefully based on these issues.

## VIII. DISCUSSION AND CONCLUSION

The objective of this paper is to produce a framework that can be a state-of-the-art technology. Therefore, this research developed an XML labelling scheme. The aim of this scheme is to improve certain functions that many existing labelling schemes suffer from. Such functions are query processing, update processing, and labelling MXL data. This paper argues that using the clustering-based techniques can improve query processing performance in the real world and labelling nodes efficiency. This technique has already been used because it has many advantages such as it decreases the re-labelling cases and therefore improves the insertion process. Improve the determination of the relationships [31]. In addition, the clustering technique has two labels for every node, and this feature improves the process of labelling nodes that from the same cluster; and the label of the cluster is used to connect that cluster with the whole tree.This feature helps in determining the relationships between the nodes that form different clusters. It is also simplify many scenarios of inserting new nodes [15]. However, despite these features of the clustering technique, it has the disadvantage that implementing two labels is a complicated process [34].

## REFERENCES


*[1] Abiteboul, S., P. Buneman, and D. Suciu, Data on the Web: from relations to semistructured data and XML. 2000: Morgan Kaufmann.*

*[2] Zhuang, C. and S. Feng. Full tree-based encoding technique for dynamic XML labeling schemes. in International Conference on Database and Expert Systems Applications. 2012. Springer.*

*[3] Bosak, J. and T. Bray, XML and the second-generation Web. Scientific American, 1999. 280(5): p. 89-93.*

*[4] Elmasri, R. and S. Navathe, Fundamentals of database systems. Vol. Global;Seventh;. 2016, Upper Saddle River: Pearson.*

*[5] Connolly, T.M. and C. Begg, Database Systems: practical approach to design, implementation, and management. 2015, Harlow: Pearson Education Limited.*

*[6] Kaushik, R., et al. Updates for structure indexes. in Proceedings of the 28th international conference on Very Large Data Bases. 2002. VLDB Endowment.*

*[7] Jiang, Y., et al., An encoding and labeling scheme based on continued fraction for dynamic XML. Journal of Software, 2011. 6(10): p. 2043-2049. Kaushik, R., et al. Covering indexes for branching path queries. in Proceedings of the 2002 ACM SIGMOD international conference on Management of data. 2002. ACM.*

*[8] Cohen, E., H. Kaplan, and T. Milo, Labeling dynamic XML trees. SIAM Journal on Computing, 2010. 39(5): p. 2048-2074.*

*[9] Tatarinov, I., et al., Storing and querying ordered XML using a relational database system. Proceedings of the ACM SIGMOD International Conference on Management of Data, 2002: p. 204-215.*
*[10] Zhang, C., et al., On supporting containment queries in relational database management systems. Proceedings of the ACM SIGMOD International Conference on Management of Data, 2001: p. 425-436.*
*[11] Li, C., et al. On reducing redundancy and improving efficiency of XML labeling schemes. in Proceedings of the 14th ACM international conference on Information and knowledge management. 2005. ACM.*
*[12] Liu, J., Z. Ma, and L. Yan. Efficient processing of twig pattern matching in fuzzy XML. in Proceedings of the 18th ACM conference on Information and knowledge management. 2009. ACM.*
*[13] O'Neil, P., et al. ORDPATHs: insert-friendly XML node labels. in Proceedings of the 2004 ACM SIGMOD international conference on Management of data. 2004. ACM.*
*[14] Xu, L., et al. DDE: from dewey to a fully dynamic XML labeling scheme. in Proceedings of the 2009 ACM SIGMOD International Conference on Management of data. 2009. ACM.*
*[15] Xu, L., T.W. Ling, and H. Wu, Labeling dynamic XML documents: an order-centric approach. IEEE transactions on knowledge and data engineering, 2012. 24(1): p. 100-113.*
*[16] Wyke, R.A. and A. Watt, XML schema essentials. 2002: Wiley.*
*[17] Harold, E.R., W.S. Means, and K. Udemadu, XML in a Nutshell. Vol. 8. 2004: O'reilly Sebastopol, CA.*
*[18] Harold, E.R. and W.S. Means, XML in a nutshell. Vol. 3rd. 2004, Beijing;Farnham;: O'Reilly.*
*[19] Garcia-Molina, H., Database systems: the complete book. 2008: Pearson Education India.*
*[20] Amato, G., et al. YAPI: Yet another path index for XML searching. in International Conference on Theory and Practice of Digital Libraries. 2003. Springer.*
*[21] Bruno, N., N. Koudas, and D. Srivastava. Holistic twig joins: optimal XML pattern matching. ACM.*
*[22] Catania, B., A. Maddalena, and A. Vakali, XML document indexes: a classification. IEEE Internet Computing, 2005. 9(5): p. 64-71.*
*[23] Rousseeuw, P.J., I. Ruts, and J.W. Tukey, The bagplot: a bivariate boxplot. The American Statistician, 1999. 53(4): p. 382-387.*
*[24] 'Neil, P., et al. ORDPATHs: insert-friendly XML node labels. 2004. ACM.*
*[25] Murata, M., S.S. Laurent, and D. Kohn, XML media types. 2000.*
*[26] Catania, B., et al. Lazy XML updates: laziness as a virtue, of update and structural join efficiency. in Proceedings of the 2005 ACM SIGMOD international conference on Management of data. 2005. ACM.*
*[27] Bruno, N.,N. Koudas, and D. Srivastava. Holistic twig joins: optimal XML pattern matching. in Proceedings of the 2002 ACM SIGMOD international conference on Management of data. 2002. ACM.*
*[28] Wang, W., et al. Efficient processing of XML path queries using the disk-based F&B index. in Proceedings of the 31st international conference on Very large data bases. 2005. VLDB Endowment.*
*[29] Klaib, A. and J. Lu. Investigation into Indexing XML Data Techniques. in Proceedings on the International Conference on Internet Computing (ICOMP). 2014. The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).*
*[30] Kaplan, H., T. Milo, and R. Shabo. A comparison of labeling schemes for ancestor queries. in Proceedings of the thirteenth annual ACM-SIAM symposium on Discrete algorithms. 2002. Society for Industrial and Applied Mathematics.*
*[31] Gusfield, D., Algorithms on strings, trees and sequences: computer science and computational biology. 1997: Cambridge university press.*
*[32] Li, Q. and B. Moon. Indexing and querying XML data for regular path expressions. in VLDB. 2001.*
*[33] Almelibari, A.A., Labelling Dynamic XML Documents: A GroupBased Approach, in Computer Science. 2015, Sheffield University: Sheffield p. 308.*

# Software Engineering Basics & the Algorithmic DocM Model for Software Development

**Azeem Uddin**
B. Tech (Computer Science & Engineering), Galgotias University, Greater Noida, Uttar Pradesh, India E-mail: uddinazeem77@gmail.com

## A B S T R A C T

*Software is a computer program that when executed provide desired features, functions & performance. Software engineering is a layered technology with quality, process, methods & tools as it's crucial layers. The development of a software requires proper documentation, analysis, planning, supervision & management. Software engineering enables us to build complex systems in a timely manner. It is the application of engineering to the development of software in a systematic method. Hence, this paper provides a general description about Software engineering & it's basics. Software is developed with the help of S DLC framework i.e. Software development lifecycle that has many phases like requirement gathering, feasibility study, system analysis, software design, coding, testing etc. S DLC is a sequence of activities that lead to the production of a software product. S o, this paper also aims to introduce a new brand S DLC Model i.e. THE ALGORITHMIC DOCUMENT MAINTENANCE MODEL. This Model is described in a thorough manner covering phases, advantages, significance & algorithmic approach.*

***Keywords - Software Engineering, Software Development Lifecycle(S DLC), Fundamentals of Software Engineering, The Algorithmic DocM Model.***

## I. INTRODUCTION

Software engineering is the combination of software & engineering. Software refers to a program or set of programs that when executed provide desired features, functions & performance whereas engineering is all about developing new products using well-defined scientific methods & principles. So, combining both the aspects, Software engineering may be defined as an engineering branch that is associated with the development of a software product. The outcome of the software development process is an efficient & reliable software product. Functionality, reliability, efficiency, usability, maintainability, portability, robustness & integrity are some of the main software characteristics. Software is developed with the help of SDLC framework i.e. Software development lifecycle that has many phases like requirement gathering, feasibility study, system analysis, software design, coding, testing etc. SDLC is a sequence of activities that lead to the production of a software product. A software consumes resources, budget & time. Development of a flexible & efficient software product requires satisfactory software project management. Software development can be divided into two parts - Software creation & Software project management. Software project management is the art & science of planning & leading software projects.

Software product is influenced by four P's i.e. People, Project, Process & Product. There are many software development lifecycle models like waterfall model, prototype model, spiral model, iterative enhancement model & many more.

Software engineering: A layered technology Software engineering is an engineering branch that is associated with the development of a software product. There are four layers that act as a base for software engineering. Due to this reason, software engineering is called as a layered technology. All the four layers of software engineering are briefly summarised below-

**1. Quality:** Quality refers to the fitness for purpose. It is a subjective aspect. There are many software quality attributes that software must possess. In fact, the bedrock that supports software engineering is quality focus.

**2. Process:** Process is one of the four P's on which software project management focuses. When a software program is executed, it becomes a process. It is a framework that must be established for effective delivery of software.

**3. Methods:** Focuses on HOW to build the software. Each method consists of multiple tasks like requirement analysis, testing, maintenance etc.

**4. Tools:** Tools are used to build up the software. Tools provide automated/semi-automated support for process & methods. For instance, CASE(Computer- aided software engineering) tools are used to design & implement applications.

## II. PRINCIPLES OF SOFTWARE ENGINEERING

Principles are basic ideas or rules that explain or control how something happens or works. Principles in software engineering are necessary for uniformity, discipline & output. Five principles of software engineering are described below-

**1. Think through the problem completely before you try to implement a solution:** This principlestates that firstly, one should understand the problem thoroughly, then only he/she should start implementing the solution because proper understanding of the problem statement & analysation process are necessary to find the appropriate solution.

**2. Divide & Conquer:** Divide & Conquer approach states that the entire problem must be divided into sub-tasks for proper management. It is also called modularization. This division of the whole work increases the reliability, concurrent execution & maintainability.

**3. Keep it simple:** It means that there is no need to make the software process complicated. Developers, project managers & the people who are associated with the software product must keep the software process simple & easy to understand & implement.

**4. What you produce, others will consume:** It is totally true that software products are made for the users or customers. Developers make them in accordance with the requirements & expectations of the users. Hence, a software must be user- friendly. User/Customer must be able to use the software with ease.

**5. Learn especially from your mistakes:** All humans make mistakes. There is nothing wrong in doing mistakes but one should learn from his/her mistakes. The same approach follows with regard to software as well. The developers must learn from their mistakes, their ultimate goal must be to produce a quality software product & they should fix the bugs/errors with great attentiveness

**The Algorithmic DocM Model for software development**

The DocM Model stands for Document Maintenance Model. This model refers to a systematic & planned approach for the development of a software. The creation of a detailed document is the most crucial aspect of this model.

## III. BASIC STEPS (Phases) INVOLVED IN THE DocM MODEL

There are many steps involved in the DocM model that must be followed sequentially & with great management. The various steps are explained below-

**1. COMMUNICATION**: The very first phase is communication phase. In this phase, the customer contacts or approaches the service provider or the developer to express his/her desire of software product. The user or customer also tries to negotiate the terms at this step.

**2. REQUIREMENT GATHERING:** At this step, requirements are gathered from the customer. It focuses on WHAT not HOW. Discussion is carried out between customer & developer. It is an activity that helps to understand what problem has to be solved & what customers expect from the software. The foundation of this phase is effective communication. There are many requirement elicitation methods

like interviews, brainstorming sessions, Facilitated application specification technique (FAST) etc. This step involves the practice of collecting the requirements from users, customers & other stakeholders.

**3. FEASIBILITY STUDY:** In this phase, it is checked that whether the project is feasible (Workable) or not. It concentrates on the operational feasibility, technical feasibility & economic feasibility. At last, a feasibility report is created that specifies whether the project is practically possible or not. This step is just an assessment of the practicality of a proposed system or project.

**4. DCT analysis phase:** Here, DCT stands for Design, Coding & Testing. This phase is not about the actual implementation, it is about the analysis & forming conclusions. The development team focuses on-

**a. DESIGN**: It is all about the modelling techniques like use case approach, Entity relationship diagram, data flow diagram etc. The developer thinks which model must be used to portray the functionalities, behaviour & structure of the software system.

**b. CODING: T**he development team decides which programming language is the most suitable for the proposed system. Actual coding is not started but a general idea is set regarding the programming methodologies. But, pseudo code or rough estimations are possible.

**c. TESTING:** Testing is the combination of verification & validation. But, at this step actual strict testing is not started. Only the test criteria & test case plan is created. More emphasis is given to the principles & objectives of testing with regard to the software project.

Hence, this DCT phase is only about analysis, planning & estimation. It is not about the actual implementation.

**5. Writing Documentation:** After the DCT analysis, there comes the role of proper documentation. A detailed document is created by the development team that comprises the following information-

a) Software requirements
b) Requirement analysis report
c) Feasibility report
d) Design details
e) Pseudo code
f) Test strategy & test plan
g) Characteristics of the software product

So, this document comprises of all the necessarydetails about the software product. The creation of this document is a time consuming process because all the details about requirements, design, coding & testing are included in it. The ultimate objective at this step is to generate a formal document that is understandable to users.

**6. Document approval or rejection phase:** After creating a final document, the developers show it to the customer for customer satisfaction & feedback. If the customer is satisfied with the document or he/she approves the document, then the developers can start the actual implementation or working from the design phase to the maintenance phase. But, if the document is not in accordance with the customer's expectations or the customer rejects the document, then a new document is created & again it will be shown to the customer until or unless the customer gets satisfied. This process will be iterated till the creation of a satisfactory document & after that it should be verified by the customer. Once it is verified by the customer, the actual working is started using required tools & methods.

**7. Delivery of software product to the customer:** Once the verification is complete, the developers try their level best to build a high quality software. When ready, this software is delivered to the customer & further maintenance is taken into account.

**ALGORITHMIC REPRESENTATION OF THE DocM MODEL**

**START**

a. Communication
b. Requirement gathering
c. Feasibility study
d. DCT analysis
e. Making of a formal Document
f. Verification of the document by the customer (User)
g. If verified, then START ACTUAL WORKING (IMPLEMENTATION) from design phase to testing phase.
h. If not verified, then make a new Document & iterate the process till the customer is satisfied.
i. Delivery of software product to the customer & maintenance.

**STOP**

**Advantages of the Algorithmic DocM Model**

**1. Algorithmic approach:** This DocM model is algorithmic in nature. It specifies the step by step procedure for the development of a software product. It focuses on input, output, effectiveness, generality & ease of use.

**2. Document-oriented:** This model is document- oriented as a detailed document is created & shown to the customer before actual implementation of the software. Documentation is the written & retained record of events that is very crucial in this world where there are fickle- minded customers. Hence, proper documentation improves the reliability, maintainability & accuracy of the software product.

**3. Customer Involvement:** Customers (Users) are the people who greatly influence the software building process. So, in this model, there is more customer involvement as after the creation of document, the document is disclosed to the customers for their feedback. It is done so as to gain the customer confidence & to check the completeness, consistency & correctness of the document.

**4. Understanding System:** Since a general idea (Outline), behaviour, functionality & structure of the software product is displayed through the easy explanation & modelling techniques, the user get a better understanding of the system.

**Significance of the Algorithmic DocM Model**

The Algorithmic DocM Model can be of utmost importance to develop a user-friendly software. This model emphasises on customer feedback & satisfaction. It is simple & easy to understand as a formal document is maintained in this process. This document plays a vital role throughout the software development cycle. It is an ALL IN ONE document that includes details regarding design, coding, testing etc. There is no need for a separate software requirement specification document. This model is Document-dependent. Written communication skills & speaking skills are used by the developers to develop the software using this algorithmic model. In this model, a document is refined again & again according to customer instructions & feedback, so this model can also handle change. After the delivery of software also, the document involved in this model is refined & maintained with operation & maintenance information.

In short, the Algorithmic DocM Model is a systematic & planned way of building a software product. the development of software in a systematic method. The power of software engineering must not be underestimated. With the help of software engineering, high quality software can be produced.

Moreover, an effective software project management can also help transfer the product from zero to hero. Hence, this paper has explored the software engineering basics & an efficacious methodology/ model for the development of robust & reliable software.

In this high tech world where user is so fickle- minded, building high quality software is not a cakewalk. But, if correct strategies & managerial aspects are followed, then quality software can surely be maintained. Even more fundamentally, it is important to recognize the need of software engineering. The Algorithmic Document Maintenance model described in the end sums this upneatly.

## IV. CONCLUSION

Software is an intangible product. The development of software requires proper documentation, analysis, planning, supervision & management. Software engineering enables us to build complex systems in a timely manner.


## ACKNOWLEDGEMENT

First & foremost, praises & thanks to the God for his shower of blessings throughout my research work to complete the research successfully.

I would like to express my deep & sincere gratitude to my parents for their love, prayers, care & sacrifices for educating & preparing me for my future. I would also like to express my gratefulness to all my academic colleagues & friends for their constant encouragement.

Finally, my special thanks go to all the people who have supported & encouraged me to complete the research work directly or indirectly.


## REFERENCES


[*1] Rajib Mall, Fundamentals of Software Engineering, PHI Publication.*
*[2] Pankaj Jalote, Software Engineering, Wiley.*
*[3] M. Cotterell, Software Project Management, Tata McGraw- Hill Publication.*

# Instructions for Authors

**Essentials for Publishing in this Journal**

1 Submitted articles should not have been previously published or be currently under consideration for publication elsewhere.

2 Conference papers may only be submitted if the paper has been completely re-written (taken to mean more than 50%) and the author has cleared any necessary permission with the copyright owner if it has been previously copyrighted.

3 All our articles are refereed through a double-blind process.

4 All authors must declare they have read and agreed to the content of the submitted article and must sign a declaration correspond to the originality of the article.

**Submission Process**

All articles for this journal must be submitted using our online submissions system. http://enrichedpub.com/ . Please use the Submit Your Article link in the Author Service area.

**Manuscript Guidelines**

The instructions to authors about the article preparation for publication in the Manuscripts are submitted online, through the e-Ur (Electronic editing) system, developed by **Enriched Publications Pvt. Ltd**. The article should contain the abstract with keywords, introduction, body, conclusion, references and the summary in English language (without heading and subheading enumeration). The article length should not exceed 16 pages of A4 paper format.

**Title**

The title should be informative. It is in both Journal's and author's best interest to use terms suitable. For indexing and word search. If there are no such terms in the title, the author is strongly advised to add a subtitle. The title should be given in English as well. The titles precede the abstract and the summary in an appropriate language.

**Letterhead Title**

The letterhead title is given at a top of each page for easier identification of article copies in an Electronic form in particular. It contains the author's surname and first name initial .article title, journal title and collation (year, volume, and issue, first and last page). The journal and article titles can be given in a shortened form.

**Author's Name**

Full name(s) of author(s) should be used. It is advisable to give the middle initial. Names are given in their original form.

**Contact Details**

The postal address or the e-mail address of the author (usually of the first one if there are more Authors) is given in the footnote at the bottom of the first page.

**Type of Articles**

Classification of articles is a duty of the editorial staff and is of special importance. Referees and the members of the editorial staff, or section editors, can propose a category, but the editor-in-chief has the sole responsibility for their classification. Journal articles are classified as follows:

**Scientific articles:**

1. Original scientific paper (giving the previously unpublished results of the author's own research based on management methods).

2. Survey paper (giving an original, detailed and critical view of a research problem or an area to which the author has made a contribution visible through his self-citation);

3. Short or preliminary communication (original management paper of full format but of a smaller extent or of a preliminary character);

4. Scientific critique or forum (discussion on a particular scientific topic, based exclusively on management argumentation) and commentaries. Exceptionally, in particular areas, a scientific paper in the Journal can be in a form of a monograph or a critical edition of scientific data (historical, archival, lexicographic, bibliographic, data survey, etc.) which were unknown or hardly accessible for scientific research.

**Professional articles:**

1. Professional paper (contribution offering experience useful for improvement of professional practice but not necessarily based on scientific methods);
2. Informative contribution (editorial, commentary, etc.);
3. Review (of a book, software, case study, scientific event, etc.)

**Language**

The article should be in English. The grammar and style of the article should be of good quality. The systematized text should be without abbreviations (except standard ones). All measurements must be in SI units. The sequence of formulae is denoted in Arabic numerals in parentheses on the right-hand side.

**Abstract and Summary**

An abstract is a concise informative presentation of the article content for fast and accurate Evaluation of its relevance. It is both in the Editorial Office's and the author's best interest for an abstract to contain terms often used for indexing and article search. The abstract describes the purpose of the study and the methods, outlines the findings and state the conclusions. A 100- to 250-Word abstract should be placed between the title and the keywords with the body text to follow. Besides an abstract are advised to have a summary in English, at the end of the article, after the Reference list. The summary should be structured and long up to 1/10 of the article length (it is more extensive than the abstract).

**Keywords**

Keywords are terms or phrases showing adequately the article content for indexing and search purposes. They should be allocated heaving in mind widely accepted international sources (index, dictionary or thesaurus), such as the Web of Science keyword list for science in general. The higher their usage frequency is the better. Up to 10 keywords immediately follow the abstract and the summary, in respective languages.

**Acknowledgements**

The name and the number of the project or programmed within which the article was realized is given in a separate note at the bottom of the first page together with the name of the institution which financially supported the project or programmed.

**Tables and Illustrations**

All the captions should be in the original language as well as in English, together with the texts in illustrations if possible. Tables are typed in the same style as the text and are denoted by numerals at the top. Photographs and drawings, placed appropriately in the text, should be clear, precise and suitable for reproduction. Drawings should be created in Word or Corel.

**Citation in the Text**

Citation in the text must be uniform. When citing references in the text, use the reference number set in square brackets from the Reference list at the end of the article.

**Footnotes**

Footnotes are given at the bottom of the page with the text they refer to. They can contain less relevant details, additional explanations or used sources (e.g. scientific material, manuals). They cannot replace the cited literature.
The article should be accompanied with a cover letter with the information about the author(s): surname, middle initial, first name, and citizen personal number, rank, title, e-mail address, and affiliation address, home address including municipality, phone number in the office and at home (or a mobile phone number). The cover letter should state the type of the article and tell which illustrations are original and which are not.

**Address of the Editorial Office:**

**Enriched Publications Pvt. Ltd.**
**S-9,**IInd FLOOR, MLU POCKET,
MANISH ABHINAV PLAZA-II, ABOVE FEDERAL BANK,
PLOT NO-5, SECTOR -5, DWARKA, NEW DELHI, INDIA-110075,
PHONE: - + (91)-(11)-45525005