Implementation of Levenshtein Distance Algorithm for E-Commerce of Bravoisitees Distro

Rusydi Umar[1], Yana Hendriana[2], Eko Budiyono[3]

1,2,3 Informatics Department, Faculty of Industrial Technology, Universitas Ahmad Dahlan, Indonesia

Abstract - *The search engine is a computer program designed to find the information sought from the amount of information that is available. By typing the word you want to search on the search engines then all the desired information is displayed. To search the possibility that the desired word, an approach requires the specific string search. In search of regular expressions, the exact search, there are various algorithms which are well known as the Knuth-Morris-Pratt, Boyer-Moore, Rabin-Karp, and others. While specific search string is that the Levenshtein Distance Algorithm approach for E-Commerce of Bravoisitees Distro. Results from this study is an e-commece web application that has its own search engine in the system. Using Levenshtein Distance algorithm, it can be performed a more accurate, even if the word is entered has a typing error then this algorithm can still find the desired data and provide search suggestions approaching from the word input.*

Keywords - *Search, Expression, Search Engine, Levenshtein Distance Algorithm.*

I. INTRODUCTION

Bravoisitees Distro located in Maguwoharjo is a fashion store that is engaged in the business of selling clothes and clothes printing. As a company that is still growing, the company continues to expand its business fields. Based on interviews conducted to the owner Bravoisitees, they stated that the number of customers who pretty much led to difficulties in the processing of data, ranging from data transactions, and the data item, because everything is still done manually with the records in the book. Actually, their customers also said that to find stuff that is sometimes quite difficult because so many goods available. While one type of goods there are only a few stocks, so that goods are hard to find. Besides marketing done by the shop is still manual, causing the turnover of income does not reach the desired target.

Based on the results of monitoring on several web e-commerce, it was found that in the case of e-commerce web search that is not able to perform accurate searches and can not find the data sought if the word entered by the user there is an error. For more details can be seen in Fig. 1 which is a search on e-commerce TokoPedia.com that displays data that does not match the input of a search word.
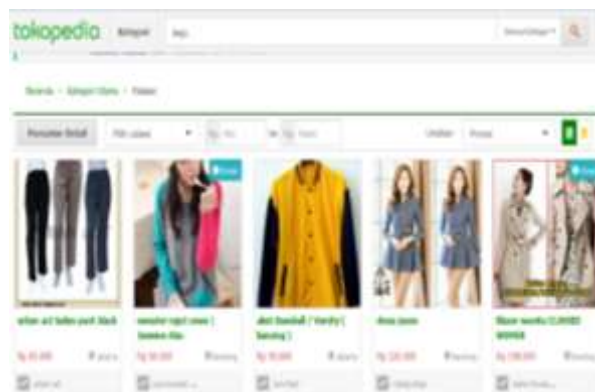

Fig. 1 E-Commerce tokopedia.com

To resolve the problems that exist between sales on Bravoisitees Distro and searching facilities problems which have not been up then this research will implements the Levenshtein Distance algorithm in the E-commerce system that can help in the process of the Searching facilities and can provide active support for the smooth business of selling goods made by Bravoisitees Distro. Online customers demonstrate that all these factors, except convenience, impact e-loyalty. The data also reveal that e-loyalty has an impact on two customer-related outcomes: word-of-mouth promotion and willingness to pay more [1].

II. LITERATURE REVIEW

Modeling and design stage includes data search, analysis, modeling and design-making system with methods of Zachman framework and models of linear sequential process. The research resulted in the modeling and design of e-commerce system with the goal after the system is built and the SMI Organization received the required information [2].

Applications *E-commerce* Sales Perfumes Online "*E-commerce* can connect sellers and buyers of different places and not be an obstacle in the transaction. Prospective buyers can find out more info about the product such as price, type, brand and so on" [3]. Implementation of *E-commerce* Sales Book Publisher Waves Based on *Framework* "The use of *e-commerce* technology provides a very broad market opportunity given the technology of the Internet has spread widely throughout the corners of the world. *E-commerce* provides many advantages and conveniences, one of which reduces operating costs and for consumers to shop faster time and a lot of information about products available" [4].

A. Levenshtein Algorithm

Phonetic distances between Standard Danish (Lyngby) and each of the other 17 Scandinavian language varieties were calculated by means of the Levenshtein algorithm. When phonetic transcriptions of two pronunciations are compared with each other, Levenshtein distance is equal to the number of operations needed to transform one transcription into another. There are three types of operations: insertions, deletions and substitutions of phonetic symbols. The power of the Levenshtein distance is that it chooses the operations that transform one pronunciation into another in such a way that the total number of string operations is minimal [5].

Number of normalized edit distances presented so far may offer good performance in some applications, none of them can be regarded as a genuine metric between strings because they do not satisfy the triangle inequality. Given two strings X and Y over a finite alphabet, this paper defines a new normalized edit distance between X and Y as a simple function of their lengths (|X| and |Y|) and the Generalized Levenshtein Distance (GLD) between them. The new distance can be easily computed through GLD with a complexity of $O(|X| \cdot |Y|)$ and it is a metric valued in [0, 1] under the condition that the weight function is a metric over the set of elementary edit operations with all costs of insertions/ deletions having the same weight. Experiments using the AESA algorithm in handwritten digit recognition show that the new distance can generally provide similar results to some other normalized edit distances and may perform slightly better if the triangle inequality is violated in a particular data set [6]. The Levenshtein edit distance allows three types of operations: an insertion, a deletion or a substitution of a character. The Damerau edit distance allows the previous three plus in addition a transposition between two adjacent characters [7]. A new technique based on the Levenshtein minimum string distance statistic for measuring error rates in text entry research. The technique obviates the need to artificially constrain subjects to maintain synchronization with the presented text, thus affording a more natural interaction style in the evaluation. Methodological implications are discussed, including the additional need to use keystrokes per characters (KSPC) as a dependent measure to capture the overhead in correcting errors [8].

B. E-Commerce

E-Commerce, is defined here as "the process of buying and selling products or services using electronic data transmission via the Internet and the www". Examples that do not fit this definition include electronic publishing to promote marketing, advertising, and customer support [10]. Most of the factors proposed as determinants of e-commerce adoption: perceived usefulness, perceived ease of use, compatibility, and external pressure were found to be statistically significant as determinants of e-commerce adoption. Reviews These results corroborate the TAM models in the sense that perceived usefulness and perceived ease of use turned out to be the most influential factors of e-commerce adoption as perceived by top managers of SMEs [10]. Electronic commerce (e-commerce) has been defined in several ways depending on the context and research objective [11,12].

E-commerce is the buying, selling and marketing goods and services through electronic systems. Many convenience gained from *e-commerce* one does not need to hold merchandise traders in the market due to the use of *e-commerce* merchants simply upload wares. Goods are uploaded can be viewed by potential buyers from different regions. Prospective buyers do not need to come to the place of the trader as examples of goods already on the web [14]. To expedite the process of search services and purchases by customers towards these online stores, we need a system that allows customers to access the service online stores, especially the process of buying fashion online using android based mobile applications store. This application will also provide facilities that assist in the search for collection until the transaction purchases by customers [13].

One of the most commonly used data mining techniques for E-commerce is finding association rules between a set of co-purchased products, dynamic set, technology, applications, and business processes that connect corporate, consumer, and certain communities through electronic transactions and trade in goods, services, and information that will be conducted electronically [15].

Recommender systems are used by E-commerce sites to suggest products to their customers. The products can be recommended based on the top overall sellers on a site, based on the demographics of the customer, or based on an analysis of the past buying behavior of the customer as a prediction for future buying behavior. Broadly, these techniques are part of personalization on a site, because they help the site adapt itself to each customer. Recommender systems automate personalization on the Web, enabling individual personalization for each customer. Personalization to this extent is one way to realize Pine's ideas on the Web. Thus, Pine would probably agree with Jeff Bezos, CEO of Amazon.com™, when he said "If I have 2 million customers on the Web, I should have 2 million stores on the Web" [16].

Organizational readiness was assessed by including two items about the financial and technological resources that the company may have available as well as factors dealing with the compatibility and consistency of e-commerce with firm's culture, values, and preferred work practices (existing technology infrastructure; and top management's enthusiasm to adopt e-commerce). Such items were found relevant in other research [17,18,19,20]. An accurate user profile can greatly improve a search engine's performance by identifying the information needs for individual users. we

proposed and evaluated several user profiling strategies. The techniques make use of click through data to extract from Web-snippets to build conceptbased user profiles automatically [21].

E-Commerce websites frequency effect on business firms have. The company's attention to its electronic gateway does not cause distortion of your account and provides self-trust can foster trade and business bankruptcy in your will. Web gateway to a brand - products and services is the product manufacturer. Website without using your colors and go like the old window and the color you used in traditional business and will cause the customers are not willing to buy the product, but also against the best site if possible through the Internet for all People must not be considered and called application (User Friendly) not, there would be no success and therefore companies should result in Internet search firm footing their own. For many users, sites search (search engine) the entry point to the Internet. However, many sites mechanism to measure trust in their portal are. Although customers are sometimes reluctant to comment or unwilling to comment on the negative, or no negative votes [22].

III. RESEARCH METHODS

A. Incremental Paradigms

The troubleshooting steps of this study in accordance with the stages of the software development process using a model that is Incremental Paradigm. Incremental model development of the system is based on a model system that is broken down so that the model of development in stages. Referring to Fig. 2, the incremental model apply the process flow linear forms of software that has not been stable over time completion of the software. Each linear sequence resulted in the addition of a piece of software in a manner similar to the addition of piecemeal (incremental) generated by a flow of the evolutionary process [9].
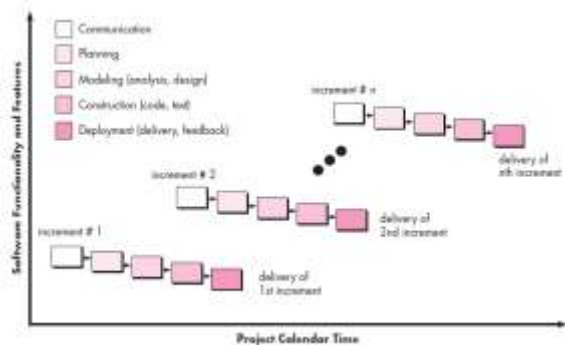

Fig. 2 Incremental Model [9]

B. Communication

Communication is the need to know the description of how the system function in accordance with the request by the owner of Bravoisitees Distro and aims to determine the needs of the users so that the system can be used and function optimally. The system will be helpful for the data searching process

and market the products owned by Bravoisitees to the public. With good publicity is expected to increase sales of Bravoisitees.

C. Planning

Planning is done for determining step arranged employment generated by the work plan within a certain time for a process of a program or application, starting from step data collection, data analysis, design, programming, implementation and testing program.

D. Modelling

This stage to define in detail the functions, constraints and objectives of software as a system specification that will be created. This stage is necessary to determine the output that will be generated by the system, the required input system, the scope of the process used to process inputs into outputs, the volume of data to be handled system, and control of the system.

E. Design

The design process will determine the overall system architecture. System design describes what should be done by the system to meet the information needs of the user. System design determines how the system will meet these objectives, consisting of design activity that generates the system specifications that meet the functional requirements developed in the process of system analysis. The design phase include:

a. Designing Data
b. Process design / Function
c. Designing User Interface

Three types of design that must be generated in the system design phase, will produce specifications for the product and the user interface method, the database structure, as well as processing and control functions in the application procedure.

F. Construction

At this stage, the software design is realized as a series of programs or program units using PHP and HTML 5. Here, the programming will be implemented gradually in accordance with the application design. Then the test will be conducted involving verification that each unit has been running the program and meet the specifications and needs to be done to test the connectivity of each software function to ensure that the system requirements have been fulfilled.

IV. RESULTS AND DISCUSSION

The system built in this research is a web e-commerce with the main function is Levenhstein Distance algorithm as a search algorithm. With the data retrieved to search based on the data items contained in the database.

The specifications of the system to be built according to user requirements, among others:

- The system can search items based on the Levenshtein Distance Algorithm.
- The system can provide suggestions to the search words.

- The system can perform data processing goods.
- The system can order goods
- Sistem dapat melakukan konfirmasi pemesanan.

A. Use Case Diagram

Use case diagrams are used to map the needs of users and admin staff as Bravoisitees. Making the use case is based on analysis of customer requirements.
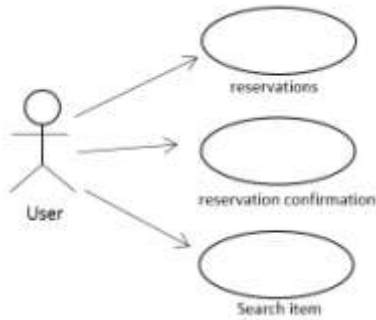

Fig. 3 Use case Diagram (user)

Fig. 3 shows that the customer is a potential buyer will get some information including a list of items that is shown is on the basis the latest goods, details of existing goods. In addition users can also use the search function to obtain the information searched goods, book and confirm reservations.

B. Activity Diagram

This activity diagram shows the steps in the work flow, decision points in the workflow, who is responsible for completing each activity, and objects used in the work flow.

1. Activity Diagram of Reservation

In the following figure is an activity diagram of the activity process of ordering goods that shows the interaction between users and systems.
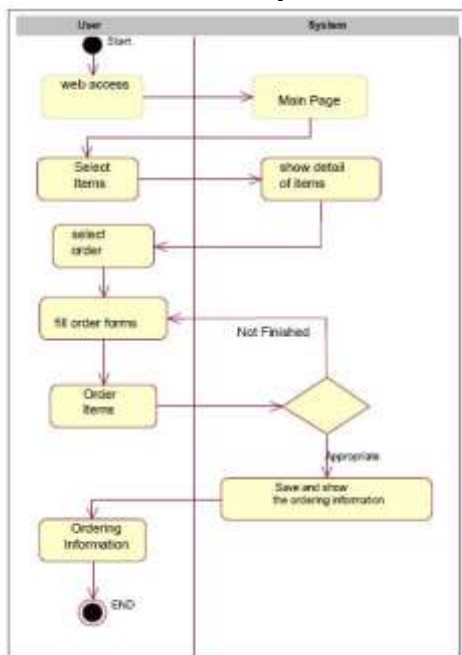

Fig. 4 Activity Diagram of Reservations

Fig. 4 shows the processes that occur in the reservation of goods ranging from accessing the web, the selection of goods, form filling up until saving reservation transaction.

2. Activity Diagram of Reservation Confirmation

The diagram in Fig. 5 shows the process of confirmation of reservation by using reservation code that has been obtained in previous reservations process, ranging from web access, charging order code, to upload of payment receipt.
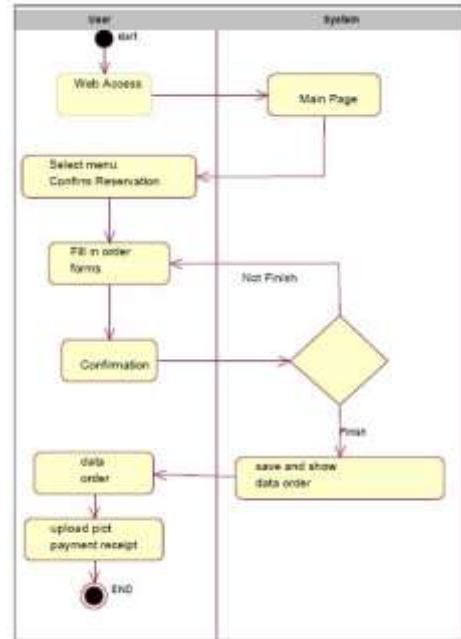

Fig. 5 Activity Diagram of Confirm Reservations

3. Activity Diagram of Searching

Activity Diagram of Searching illustrates the process of searching by keyword entered by user who then keyword will be matched using the Levenshtein Distance Algorithm, then matching results will be displayed as search results to the user, as shown in Fig. 6.
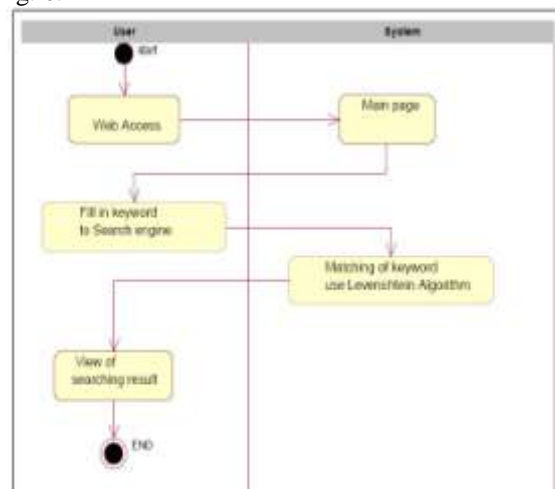

Fig. 6 Activity Diagram of Searching

C. Design Implementation

Implementation used to implement the draft design of the user interface (UI) system has become an e-commerce application based on the analysis of the system described in the previous stage. This stage is the process of computerization of the system software that has been generated in studies using PHP with the framework CodeIgneter.

1. Main Page

Main page showing the goods sold in Bravoisitees Distro and also menus required by the customer, as shown in fig. 7.
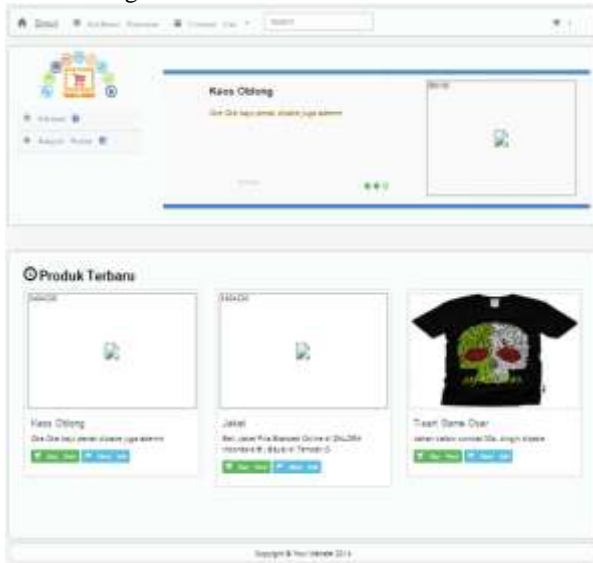
Fig. 7. Main Page

2. Reservation Page

Customers can order goods by filling identification, and after customers make reservation then the customer will receive reservation code that is used to process the registration and displays the details of the goods ordered, as shown in fig. 8 and fig. 9.

Fig. 8 Form Fill in of Customer's Identity

Fig. 9 Details of items ordered

3. Confirmation Form Page

Customers can confirm payment for goods that have been ordered by filling a purchase code that has been gained in the process of ordering goods, then the customer is required to submit pictures of payment receipt to confirm, as shown in fig. 10.
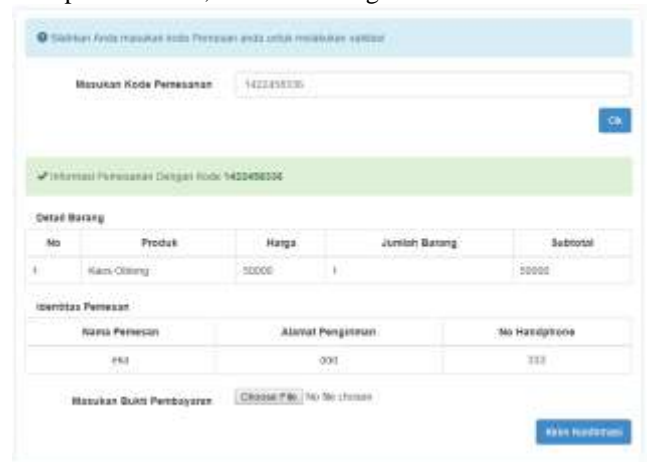
Fig. 10 Confirmation Form Page

4. Searching Page

On this search form Levenshtein Distance algorithm is applied, which aims to facilitate the search data by checking the word of the keywords entered by the customer and then be made to match the name of the items in the system, as shown in fig. 11.
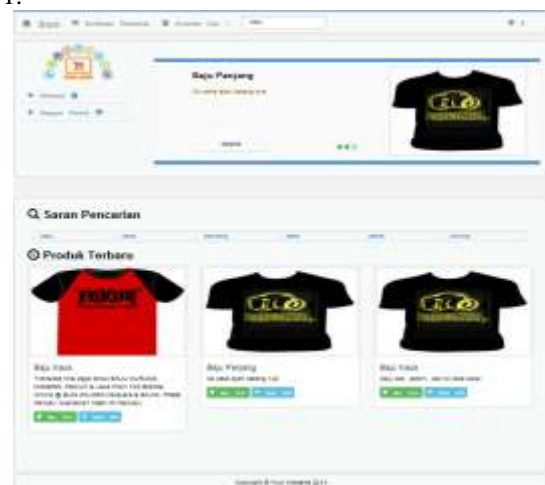
Fig. 11 Searching Page

V. CONCLUSION

The conclusions of this research are as follows:

- The establishment of system E-Commere with a search system Levenshtein Distance.
- Testing expediency concept Levenshtein Distance algorithm which can be determined based on the effective data search results even though there is so much data stored in the system.
- Search the number of operations in the approximate string matching string needed to transform one string into another string require exact algorithm like Levenshtein Distance algorithm.

REFERENCES

[1] Srinivasan, SS, Anderson, R., & Ponnavolu, K. 2002. Customer loyalty in e-commerce: an exploration of its antecedents and consequences. Journal of Retailing, 78 (1), 41-50.

[2] Hendriana, Y., Umar, R., Pranolo, A. 2015. Modelling and Design E-Commerce SMI Sector Using Zachman Framework. International Journal of Computer Science and Information Security, 13(8), 9-14.

[3] Fansyuri, Ahmad. 2012. Aplikasi E-Commerce Penjualan Parfum Secara Online. Skripsi. Yogyakarta: Program Studi Teknik Informatika Universitas Ahmad Dahlan.

[4] Yanti, Nur Fitri. 2011. Implementation of E-Commerce Sales Book on Ombak Publisher Based Framework. Skripsi. Yogyakarta: Informatics Department Universitas Ahmad Dahlan.

[5] Beijering, K., Gooskens, C., & Heeringa, W. 2008. Predicting intelligibility and perceived linguistic distances by means of the Levenshtein algorithm.Linguistics in the Netherlands, 15, 13-24.

[6] Yujian, L., & Bo, L. 2007. A normalized Levenshtein distance metric. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 29(6), 1091-1095.

[7] Hyyrö, H. 2003. A bit-vector algorithm for computing and Damerau Levenshtein edit distances. Nord. J. Comput., 10 (1), 29-39.

[8] Soukoreff, RW, & MacKenzie, IS. 2001. Measuring errors in text entry tasks: an application of the Levenshtein string distance statistic. In CHI'01 extended abstracts on Human factors in computing systems (pp. 319-320). ACM.

[9] Basuki Achmad, 2011, Software Process Model, Jakarta, PT. Mizan Publika.

[10] Grandon, E. E., & Pearson, J. M. 2004. Electronic commerce adoption: an empirical study of small and medium US businesses. Information & management, 42(1), 197-216.

[11] E.W.T. Ngai, F.K.T. Wat,. 2002. A literature review and classification of electronic commerce research, Information and Management 39, pp. 415-429.

[12] G.P. Schneider, J.T. Perry. 2000. Electronic Commerce, Course Technology, Cambridge, MA.

[13] Hendriana, Y., Hardi, R., & Pranolo, A. 2015. Design and Implementation of Online Fashion Store "Demi Outfits" Based on Android. International Journal of Computer Applications Technology and Research, Volume 4–Issue 6, 438 - 443

[14] Wong, Jony. 2010. Internet marketing for beginners. Elex Media Komputindo.

[15] Sarwar, Badrul, et al. 2000."Analysis of recommendation algorithms for e-commerce." Proceedings of the 2nd ACM conference on Electronic commerce. ACM,

[16] Schafer, J. Ben, Joseph Konstan, and John Riedl. 1999. "Recommender systems in e-commerce." Proceedings of the 1st ACM conference on Electronic commerce. ACM.

[17] R.C. Beatty, J.P. Shim, M.C. Jones. 2001. Factors influencing corporate web site adoption: a time-based assessment, Information and Management 38, pp. 337–354.

[18] W.W. Chin, A. Gopal. 1995. Adoption intention in GSS: relative importance of beliefs, DATA BASE 26 (2–3), pp. 42–64.

[19] G. Premkumar, M. Potter. 1995. Adoption of computer aided software engineering (CASE) technology: an innovation adoption perspective, DATA BASE 26 (2–3), pp. 105–123.

[20] J.Y.L. Thong. 2001. Resource constraints and information systems implementation in Singaporean small businesses, OMEGA 29, pp. 143–156.

[21] Sivaraman. K. 2011. Collaborative Filtering Based On Search Engine Logs, International Journal of Computer Trends and Technology (IJCTT) – volume 2 Issue 2 Number 4, pp. 1-6.

[22] Najafi, I. 2014. Identify Effective Factors for Improving E-Trust of E-Transactions in the Context of E-Commerce and E-Government, International Journal of Computer Trends and Technology (IJCTT) – volume 17 Number 6, pp. 281-299.