Investigation of Disease from Multiple Health Care Data by Using HL7 Message on HDFS

M. Ramamoorthy

Abstract: To furnish a detailed analysis on the healthcare data in HL7(Health Level 7) by utilizing Big data analytics and to create disease prediction system as the healthcare sector is considered as one of the important industries in information technology. In the past few years, healthcare data has become more complex with huge amount of data related to ever growing and changing technologies, mobile applications and discovery of new diseases. The belief of healthcare sectors is that healthcare data analytics tools are important to manage a large amount of complex data, to improve healthcare industries, efficiency and accuracy in medical practice. Information technology is a boon for the healthcare sectors to upgrade their efficiency in healthcare performance availing the provided data and information. The concept of big data though not new is constantly changing with attempts of defining big data and inventing newer hardware and software mechanisms to store, analyze and visualize the data in keeping with the requirements of collection of data elements as per the requisite size, speed and type. Velocity, variety and volume are the three innate aspects of the data produced. Furthermore, each of these data repositories is isolated and inherently incapable of providing a platform for global data transparency.

Keywords: Health Level 7 (HL7), Healthcare, Hadoop Distributed File System (HDFS), Deep Analysis.

INTRODUCTION

Healthcare is no exception in the field of big data. Big Data technologies are changing the world. Hospitals have to follow the changing technologies to upgrade, to encounter and to solve Big Data problems that include collecting, processing, storing, analyzing, and retrieving the real-time and accumulated healthcare data using relevant Big Data technologies. The analytical data management applications is altering speedily and the necessity of ever mounting amount of data to be stored and provided by analytical database system is enormous due to increased automation with which data can be produced. Thereby errors are prevalent as more nodes has to be added to the system. With ever increasing growth of data to be analyzed and hundreds of nodes are multiplying Clinical systems which mainly use relational databases management systems for storing transactional structured, and semi- or unstructured data in RDB tables.

Healthcare providers such as hospitals and labs contain data systems that are composed of multiple, distributed applications.

HL7 is a data exchange messaging standard for moving clinical and administrative information among healthcare applications. Healthcare providers have many different computer systems used for storing patient records. All of these systems should communicate with each other when they receive new information, or when they want to retrieve information. HL7 messages contain the processes which include admitting a patient to a hospital or requesting a lab order for a blood test. Message profiles are used to limit the set of data that is exchanged among HL7 message that provide the requirements that can verify relevant data exchange. To provide an analysis of disease using the data from multiple health-care data sources, the data has to be converted into a particular and specific format which is HL7.

Deep analysis is a process that is used for analysis and extraction of information that are organized as a very large volume of data in a form which is acceptable, useful and beneficial for an organization.

M. Ramamoorthy, Assistant Professor, Department of Computer Science and Engineering, BIST, BIHER, Bharath Institute of Higher Education & Research, Selaiyur, Chennai. E-mail: saimoorthy123@gmail.com

Deep analysis retrieves targeted information from data stores through data processing techniques. The concept of deep analysis is that it allows elimination of the unwanted data.

RELATED WORK

Efforts are being continued to merge ideas that are collected from the MapReduce and database systems. This work focuses mainly on language and interface issues. The protocol used for the communication between the Electronic Health Record components is called as the HL7. It contains system architecture which is open that entitles interfacing between systems using relevant protocols, independent of the vendor.

Both present and future is benefitted by following a standard protocol to get connected to any system that supports this particular part of the standard.

Interface allows for numerous systems to be added to a single HL7 feed. Not necessary to modify the original source code while adding the new system. The main drawback of HL7 is not really plug-and-play. HL7 implementations differ between vendors. Due to missing fields and values, duplication of data in fields and formats and different versions exist. A major disadvantage is that most of them are controlled by EHR vendors and therefore interoperability of data is a bit of challenge.

BACKGROUND

A. Hadoop Distributed File System (HDFS)

Hadoop provides a distributed file system and a framework for the analysis of very huge data sets using the MapReduce algorithm. Hadoop is mainly used for the splitting the data and computation across huge volume of hosts and also the execution of the application is parallely computed to their data. A Hadoop cluster is used for the computation capacity, storage capacity and I/O bandwidth by simply adding commodity servers.

Hadoop Distributed File System (figure 1) stores files system metadata and the application data separately. HDFS will store the metadata on a dedicated server which is called the Name Node. The Data Nodes contains the application data. All the servers are fully connected and it will communicate with each other by using the TCP-based protocols

B. Mapreduce

MapReduce is a processing mechanism. MapReduce is a program model for distributed computing based on java. The MapReduce algorithm contains two main tasks that are Map and Reduce.

During the Map phase it takes a set of data as the input data, it is divided into input splits for analysis by map task running in parallel and converts it into another set of data, where individual elements are broken down into tuples namely the key/value pairs.

Secondly, the reduce phase which takes the output from the map phase as an input and it combines the data tuples into a minimum set of tuples. The Reduce phase is used to reduce the task and provide the final result. As the sequence of the MapReduce, the reduce task is always performed after the map job is completed. The main advantage of MapReduce is that it is easy for data processing over many computing nodes.

PROPOSED SYSTEM

The clinical data from the daily operations of MC hospitals and clinics are made up of a variety of document types – many in the format of HL7 V2 messages. Creating a new patient record or updating an existing patient record may result in the creation of one or more HL7 messages. Each document type is managed or generated by a single source system at MC. We develop the application for deep analyze in the healthcare data HL7 message format.

Proposed work not process only inside the hospital this system mainly focused on disease prediction for various strategies, this source data started from medical laboratory normally users are going to check their health condition in diagnostic center.

They transfer all the patient disease information to healthcare care officer, the user also can provide their medical info to the health care officer, they collect all the diagnostic center information all well as the patient each and every area this medical report forward to the government medical research center. Medical reports convert to the HL7 message by using HAPI HL7 v2 format stored in the HDFS. Then the research center can analyze the healthcare information from various area and city.

SYSTEM ARCHITECTURE
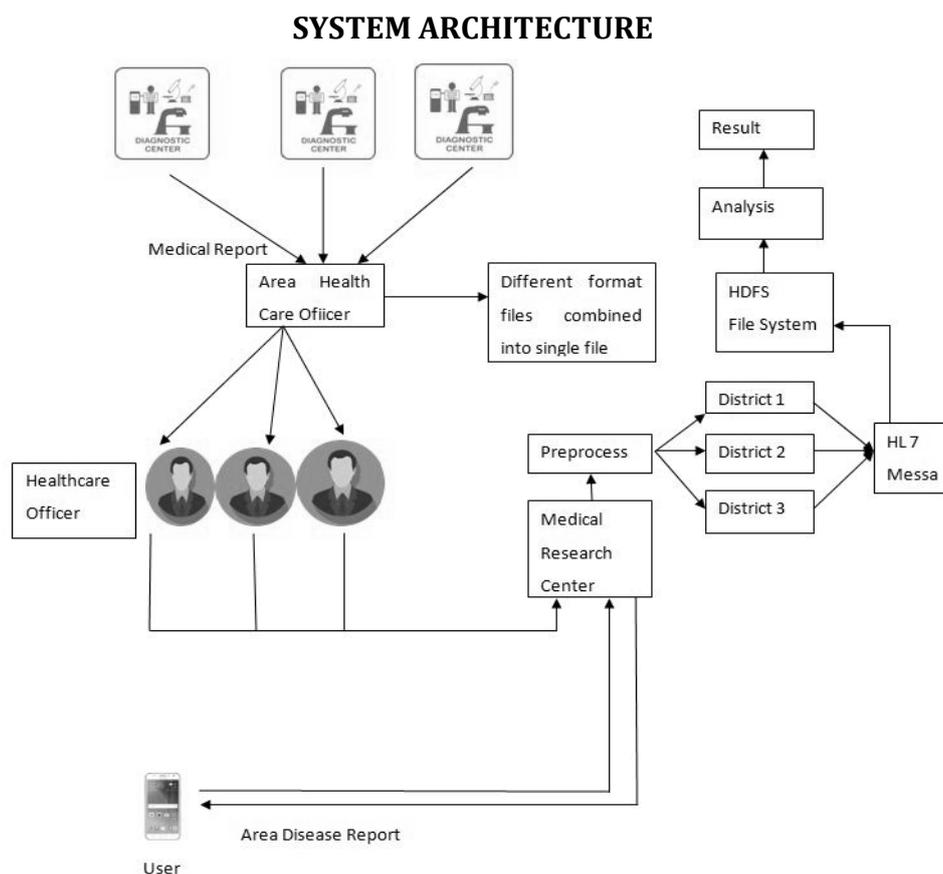
Fig.1: System Architecture

MODULES DESCRIPTION AND IMPLEMENTATION

The figure 1 is explained below.

A. Diagnostic Center Report Submission:

Every hospital having diagnostic center some places diagnostic center working separately. Doctors will suggest diagnostic center to the patient they want to know about patient health condition and disease. After taking medical checkup in the diagnostic center this data will be transferred to the District Healthcare officer this same process will be applicable for the every diagnostic center. But the transfer medical report not contains any personal information about the patient.

B. District Healthcare Officer and User

The healthcare officer getting information from every diagnostic center, all the healthcare care report will be combined into a single file. The user can update their disease information directly to health care officer, after submission the user will get notification about the disease and precautions. Healthcare officer transferred entire detail to the medical research center.

C. HL7 Message Preprocessing

Healthcare officer report will be converted into HL7 message format and stored into the HDFS (Hadoop Distributed File System).The preprocessing is a kind job to process the HL7 messages, where the level of process happening in different procedure, the report contains all medical information will be categorized by age, gender, symptoms and so on, this preprocessing is triggered at times automatically in our application. So the medical analysis information dynamically changed every time.

D. Medical Research Center Analysis

This is the Government medical research center analyze the medical report in various category. They can find out which disease affected most of the area and many people, if suppose result displayed disease dengue system should give detail about the disease. Which aged peoples are affected by the dengue, which symptoms peoples are affected by this disease which area people affected by this disease all these details will be analyzed in this application. Result showing any district or area the precautions will be delivered to those users.

RESULT
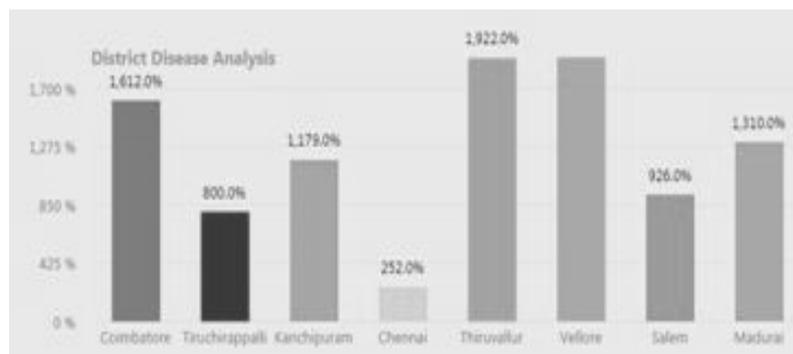
Fig.2: District-wise Disease Analysis

Invention of disease prediction system graphs out minute details of specific diseases, area spread, viral diseases, symptoms, treatment healthcare facilitates available and the necessary precaution to be undertaken to keep one from being affected.
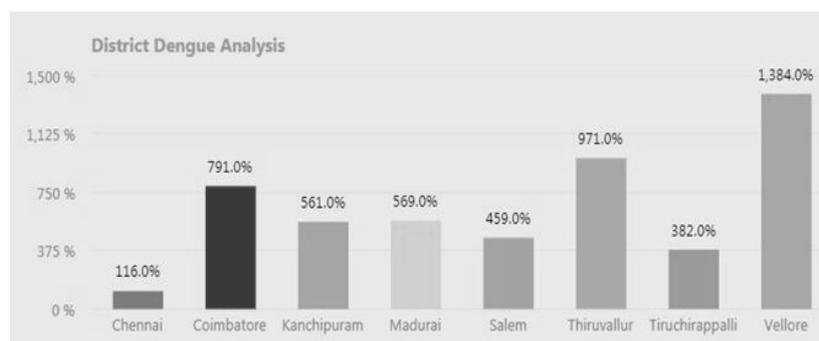
Fig.3: District-wise Dengue Analysis

End users has to download the mobile application, register and log in to view and understand the complete details of spread of specific diseases, area prone to disease, symptoms, risk age group, details of treatment and also to educate on the necessary precaution to be undertaken to safeguard oneself and thus create an awareness among public. Using deep analytics the district disease analysis (Figure 3) is calculated and the depth first search is used to analyze the district dengue analysis (Figure 4). It is colorfully illustrated in graphic method to simplify for the better understanding, thereby, enhance the civic sense of common man. In today's fast paced world, not only we see advancement in new technology but also the rapid surge of new diseases.

CONCLUSION AND FUTURE WORK

This project gives the amount of people affected by the disease which is spreading in that particular area or city .Using deep analysis technique the spreading diseases are calculated and analyzed based on the area. This helps to take precaution as well as to protect themselves from spreading diseases. This work is presented to mobile user to make them easily understand the amount of people got affected in their area. This platform needs to get strengthened in the near future by correctly implementation of the following policies:

1) Evaluate the extent of heavy spread of diseases.
2) Educate the necessary precaution to be undertaken.
3) Recommendation to the government to utilize the application which can figure out the onset of disease

REFERENCES

[1]   Udayakumar, R., Khanaa, V., & Saravanan, T. (2013). Analysis of polarization mode dispersion in fibers and its mitigation using an optical compensation technique. *Indian Journal of Science and Technology*, *6*(6), 4767-4771.

[2]   Udayakumar, R., Kumaravel, A., & Rangarajan, K. (2013). Introducing an efficient programming paradigm for object-oriented distributed systems. *Indian Journal of Science and Technology*, *6*(5S), 4596-4603.

[3] Mageswaran, S.U., & Sekhar, N.G. (2013). Reactive power contribution of multiple STATCOM using particle swarm optimization. *International Journal of Engineering & Technology*, *5*(1), 122-126.

[4] Giri, R.K., & Saikia, M. (2013). Multipath routing for admission control and load balancing in wireless mesh networks. *International Review on Computers and Software*, *8*(3), 779-785.

[5] Padmapriya, G., Manikandan, A., Krishnasamy, V., Jaganathan, S.K., & Antony, S.A. (2016). Spinel NixZn1− xFe2O4 (0.0≤ x≤ 1.0) nano-photocatalysts: synthesis, characterization and photocatalytic degradation of methylene blue dye. *Journal of Molecular Structure*, *1119*, 39-47.

[6] Vijayaragavan, S.P., Karthik, B., Kiran Kumar, T.V.U., & Sundar Raj, M. (2013). Analysis of chaotic DC-DC converter using wavelet transform. *Middle-East Journal of Scientific Research*, *16*(12), 1813-1819.

[7] Lokesh, K., Kavitha, G., Manikandan, E., Mani, G.K., Kaviyarasu, K., Rayappan, J.B.B., ... & Maaza, M. (2016). Effective ammonia detection using n-ZnO/p-NiO heterostructured nanofibers. *IEEE Sensors Journal*, *16*(8), 2477-2483.

[8] Abraham, A.G., Manikandan, A., Manikandan, E., Vadivel, S., Jaganathan, S.K., Baykal, A., & Renganathan, P.S. (2018). Enhanced magneto-optical and photo-catalytic properties of transition metal cobalt (Co2+ ions) doped spinel MgFe2O4 ferrite nanocomposites. *Journal of Magnetism and Magnetic Materials*, *452*, 380-388.

[9] Kennedy, J., Fang, F., Futter, J., Leveneur, J., Murmu, P.P., Panin, G.N., & Manikandan, E. (2017). Synthesis and enhanced field emission of zinc oxide incorporated carbon nanotubes. *Diamond and Related Materials*, *71*, 79-84.

[10] Teresita, V.M., Manikandan, A., Josephine, B.A., Sujatha, S., & Antony, S.A. (2016). Electromagnetic properties and humidity-sensing studies of magnetically recoverable LaMg x Fe 1− x O 3− δ perovskites nano-photocatalysts by sol-gel route. *Journal of Superconductivity and Novel Magnetism*, *29*(6), 1691-1701.

[11] Caroline, M.L., & Vasudevan, S. (2009). Growth and characterization of pure and doped bis thiourea zinc acetate: Semiorganic nonlinear optical single crystals. *Current applied physics*, *9*(5), 1054-1061.

[12] Jayalakshmi, V., & Gunasekar, N.O. (2013). Implementation of discrete PWM control scheme on Dynamic Voltage Restorer for the mitigation of voltage sag/swell. *International Conference on Energy Efficient Technologies for Sustainability*, 1036-1040.

[13] Udayakumar, R., Khanaa, V., & Kaliyamurthie, K.P. (2013). Optical ring architecture performance evaluation using ordinary receiver. *Indian Journal of Science and Technology*, *6*(6), 4742-4747.

[14] Udayakumar, R., Khanaa, V., & Kaliyamurthie, K.P. (2013). Performance analysis of resilient ftth architecture with protection mechanism. *Indian Journal of Science and Technology*, *6*(6), 4737-4741.

[15] Saravanan, T., Srinivasan, V., & Sandiya, V.P. (2013). A two stage DC-DC converter with isolation for renewable energy applications. *Indian Journal of Science and Technology*, *6*(6), 4824-4830.

[16] Sundarraj, M. (2013). Study of compact ventilator. *Middle-East Journal of Scientific Research*, *16*(12), 1741-1743.

[17] Thema, F.T., Manikandan, E., Gurib-Fakim, A., & Maaza, M. (2016). Single phase Bunsenite NiO nanoparticles green synthesis by Agathosma betulina natural extract. *Journal of alloys and compounds*, *657*, 655-661.

[18] Sathyaseelan, B., Manikandan, E., Sivakumar, K., Kennedy, J., & Maaza, M. (2015). Enhanced visible photoluminescent and structural properties of ZnO/KIT-6 nanoporous materials for white light emitting diode (w-LED) application. *Journal of Alloys and Compounds*, *651*, 479-482.

[19] Gopalakrishnan, K., Prem Jeya Kumar, M., Sundeep Aanand, J., & Udayakumar, R. (2013). Analysis of static and dynamic load on hydrostatic bearing with variable viscosity and pressure. *Indian Journal of Science and Technology,* 6(6), 4783-4788.

[20] Prabhu, M.R., Reji, V., & Sivabalan, A. (2012). Improved radiation and bandwidth of triangular and star patch antenna. *Research Journal of Applied Sciences, Engineering and Technology*, *4*(12), 1740-1747.

[21] Arumugam, S. and Ramareddy, S. (2012). Simulation comparison of class D/ Class E inverter fed induction heating. *Journal of Electrical Engineering*, 12(2), 71-76.

[22] Udayakumar, R., Khanaa, V., & Kaliyamurthie, K.P. (2013). High data rate for coherent optical wired communication using DSP. *Indian Journal of Science and Technology*, *6*(6), 4772-4776.

[23] Nagarajan, C., & Madheswaran, M. (2012). Experimental Study and Steady State Stability Analysis of CLL-T Series Parallel Resonant Converter with Fuzzy Controller using State Space Analysis. *Iranian Journal of Electrical and Electronic Engineering*, *8*(3): 259-267.

[24] Gopalakrishnan, K., PremJeya Kumar, M., SundeepAanand, J., & Udayakumar, R. (2013). Thermal properties of doped azopolyester and its application. *Indian Journal of Science and Technology*, *6*(6), 4722-4725.

[25] Kumaravel A., Meetei O.N. (2013). An application of non-uniform cellular automata for efficient cryptography. *Indian Journal of Science and Technology*, *6*(5): 4560-4566.

[26] Kumaravel, A., & Pradeepa, R. (2013). Layered approach for predicting protein subcellular localization in yeast microarray data. *Indian Journal of Science and Technology*, *6*(5S), 4567-4571.

[27] Kaviyarasu, K., Manikandan, E., Kennedy, J., & Maaza, M. (2016). Synthesis and analytical applications of photoluminescent carbon nanosheet by exfoliation of graphite oxide without purification. *Journal of Materials Science: Materials in Electronics*, *27*(12), 13080-13085.

[28] Mathubala, G., Manikandan, A., Antony, S.A., & Ramar, P. (2016). Photocatalytic degradation of methylene blue dye and magneto-optical studies of magnetically recyclable spinel NixMn1-xFe2O4 (x= 0.0–1.0) nanoparticles. *Journal of Molecular Structure*, *1113*, 79-87.

[29] Manikandan, E., Kennedy, J., Kavitha, G., Kaviyarasu, K., Maaza, M., Panigrahi, B.K., & Mudali, U.K. (2015). Hybrid nanostructured thin-films by PLD for enhanced field emission performance for radiation micro-nano dosimetry applications. *Journal of Alloys and Compounds*, *647*, 141-145.

[30] Kumaravel, A., & Meetei, O.N. (2013). An application of non-uniform cellular automata for efficient cryptography. *IEEE Conference on Information & Communication Technologies*: 1200-1205.

[31] Elijah, Aiden, (2018). Co-Clustering based Cross-Domain Text Classification Algorithm with Semantic ANALYSIS for Wikipedia.*Journal of Computational Information Systems, 14*(3), 44 - 49.

[32] Asim, M., Gopalia, R., & Swar, S.(2014). Comparison of Methods for Solving Travelling Salesmen Problem.*International Journal of Advances in Engineering and Emerging Technology, 5*(2), 80-87.

[33] Kowsalya, P., & Rajhitha, G. (2015). Design and Implementation of AMI System for High Traffic Smart Grid Applications using ZigBee. *Excel International Journal of Technology, Engineering and Management, 2*(1), 14-18.

[34] Skaria, B., Dr. John, E.T., & Shajan, P.X. (2016). Literature Review on Web Mining. *Bonfring International Journal of Data Mining, 6*(1), 04-06.

[35] Moravej, Z., Behravesh, V., & Bagheri, S. (2015). Optimal PMU Placement for Power System Using Binary Cuckoo Search Algorithm. *International Academic Journal of Innovative Research, 2*(10), 8-19.

[36] Retheesh, D. (2014). Analysis on FPGA Designs of Parallel High Performance Multipliers. *International Journal of Communication and Computer Technologies, 2*(1), 11-18.

[37] Khan, A., & Dr. Khushnood, S. (2017). Simple and Efficient Blood Glucose Measurement Technique Using Non Invasive Artificial Intelligence. *Bonfring International Journal of Industrial Engineering and Management Science, 7*(1), 09-13.

[38] Omidvar, R. (2015). OOSP (Off-On Sources Problem). *International Academic Journal of Science and Engineering, 2*(5), 22-31.

[39] Rajalakshmi, J., (2014). Implantable CPW Fed X-Shaped Monopole Antenna for Biomedical Application. *International Journal of System Design and Information Processing, 2*(1), 23-26.

[40] Kaveen, P., & Dr.Singaravel, G. (2018). Simulation of Efficient Life-Time in Clustering Approaches for New Approach in Wireless Sensor Network. *Bonfring International Journal of Networking Technologies and Applications, 5*(1), 1-2.

[41] Langeswaran, K., Gowthamkumar, S., Vijayaprakash, S., Revathy, R., & Balasubramanian, M.P. (2013). Influence of limonin on Wnt signalling molecule in HepG2 cell lines. *Journal of natural science, biology, and medicine*, *4*(1), 126-133.

[42] Srinivasan, V., & Saravanan, T. (2013). Analysis of harmonic at educational division using CA 8332. *Middle-East Journal of Scientific Research*, *16*(12), 1768-73.

[43]   Josephine, B.A., Manikandan, A., Teresita, V.M., & Antony, S A. (2016). Fundamental study of LaMg x Cr 1– x O 3– δ perovskites nano-photocatalysts: sol-gel synthesis, characterization and humidity sensing. *Korean Journal of Chemical Engineering*, *33*(5), 1590-1598.

[44]   Saravanan, T., Saritha, G., & Udayakumar, R. (2013). Robust H-Infinity Two Degree of Freedom Control for Electro Magnetic Suspension System. *Middle-East Journal of Scientific Research*, *18*(12), 1827-1831.

[45]   Rajasulochana, P., Dhamotharan, R., Murugakoothan, P., Murugesan, S., & Krishnamoorthy, P. (2010). Biosynthesis and characterization of gold nanoparticles using the alga Kappaphycus alvarezii. *International Journal of Nanoscience*, *9*(05), 511-516.

[46]   Slimani, Y., Güngüneş, H., Nawaz, M., Manikandan, A., El Sayed, H. S., Almessiere, M. A., & Baykal, A. (2018). Magneto-optical and microstructural properties of spinel cubic copper ferrites with Li-Al co-substitution. *Ceramics International*, *44*(12), 14242-14250.

[47]   Kaviyarasu, K., Manikandan, E., Kennedy, J., Jayachandran, M., & Maaza, M. (2016). Rice husks as a sustainable source of high quality nanostructured silica for high performance Li-ion battery requital by sol-gel method–a review. *Adv. Mater. Lett*, *7*(9), 684-696.

[48]   Ilayaraja, K., & Ambica, A. (2015). Spatial distribution of groundwater quality between injambakkamthiruvanmyiur areas, south east coast of India. *Nature Environment and Pollution Technology*, *14*(4), 771-776, 2015.

[49]   Sharmila, S., Rebecca, L. J., Das, M.P., & Saduzzaman, M. (2012). Isolation and partial purification of protease from plant leaves. *Journal of Chemical and Pharmaceutical Research*, *4*(8), 3808-3812.

[50]   Rajakumari, S.B., & Nalini, C. (2014). An efficient cost model for data storage with horizontal layout in the cloud. *Indian Journal of Science and Technology*, *7*(3), 45-46.