An Open Access Journal available online

Covenant Journal of Informatics & Communication Technology (CJICT)

Vol. 4 No. 2, Dec, 2016

A Bi-annual Publication of the Departments of Computer Information Science, and Electrical & Information Engineering. Covenant University, Canaan Land, Km 10, Idiroko Road, Ota, Ogun State, Nigeria.

Editor-in-Chief: Prof. Sanjay Misra
sanjay.misra@covenantuniversity.edu.ng

Managing Editor: Edwin O. Agbaike
me@covenantuniversity.edu.ng

Website: http//Journal.covenantuniversity.edu.ng/cjict/

© 2016, Covenant University Journals

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any meams, electronic, electrostatic, magnetic tape, mechanical, photocopying, recording or otherwise, without the prior written permission of the publisher.

It is a condition of publication in this journal that manuscripts have not been published or submitted for publication and will not be submitted or published elsewhere.

Upon the acceptance of articles to be published in this journal,the author(s) are required to transfer copyright of the article to the publisher.

ISSN - Print: 2354 – 3566
- Electronics: 2354 – 3507

Published by Covenant University Journals,
Covenant University, Canaanland, Km 10, Idiroko Road,
P.M.B. 1023, Ota, Ogun State, Nigeria

Printed by Covenant University Press

Articles

Covenant Journal of Informatics & Communication Technology Vol. 4 No. 2, Dec, 2016

An Open Access Journal, Available Online

Data Mining Approach for Target Marketing SMEs in Nigeria

Afolabi Ibukun.T., Rowland E. Worlu and Uwadia C.O.

[1]Department of Computer and Information sciences,
Covenant University, Ota, Nigeria
{ibukun.fatudimu, rowland.worlu@covenantuniverity.edu.ng

Abstract: Presently, in addition to the lack of funds experienced by SMEs in Nigeria, the main challenges experienced by SMEs (Small and Medium Enterprises) in Nigeria has to do with them not being able to understand and the apply marketing concept. Also there are a lot of resources being wasted in marketing campaign which does not target anybody in particular. This study is therefore focused on finding the target market of SMEs using the K-means clustering technique in combination with sentiment analysis, also known as opinion mining. The study also aims at recommending the best medium to market these SMEs as discovered from the output of the analysis. The result of this research will give positive direction to improving the profit of small and medium business through target marketing. Also, this study is not trivial because it will help to reduce marketing cost through target marketing. The study also helps to discover the opinion of the populace on the small and medium business and the medium through which they can be targeted. This discovery will definitely form solid foundation for further marketing action and improve their profit.

Keywords: Data Mining, K-means Clustering, Target Marketing, Sentiment Analysis, SME, Opinion Mining,

1. Introduction

According to Gulani and Aisha (2012), SMEs has no universally accepted definition, the grouping of enterprise as small, medium or large, is different from country to country. In Nigeria, in 2004, Chukwemeke defined a small scale business as one which has a total asset including capital, equipment, plant and working capital as less than N250,000. Such business is said to employ fewer than 50 full time workers. SME was defined by the Central Bank of Nigeria (CBN, 2011) as an enterprise whose

annual turnover is between N25, 000-N50, 000. NIDB (Nigerian Industrial Development Bank) is also noted to have defined Small business as enterprises that their investment and working capital is not more than N750,000. On the other hand it defined as medium business, such that have projects costs that ranges between N750,000 to N3 million (Ogbadu, 2012). According to (Pusatli and Misra, 2011), SMEs (involved in software development) can be described to have financial and employee Constraints.

According to (Eniola, 2014), SMEs have taken a leading role in the growth of the Nigerian economy; it is the driving force and establishes an important mainstay of the Nigerian economy. Also, SMEs has the ability to provide employment to the populace, improve local technology, diversify output, develop indigenous entrepreneurship and provide forward integration with large scale industries (CBN, 2011b). Despite this claim, SMEs has not been able to meet up to expectation (Osotimehin et al., 2012). Research also reveals that surviving challenges of SME managers tend to overshadow the importance of marketing (Scheers, 2010).

Nigeria SMEs have not done so well despite the fact that it constitute more than 90% of Nigerian businesses, its contribution to the nation's GDP is below 10% (Gbandi and Amissah, 2014). This could be attributed to poor performance of SMEs as regards making profits, through customer patronage amongst others such as lack of adequate financing and so on. This study is therefore aiming to help such business make more profit by improving their customer base through target marketing. In this research paper, we refer to target marketing as the ability to segregate a market into smaller groups. These groups of buyers will have clear needs, characteristics, or behaviors who might require different products or marketing mixes." (Sulekha, 2011). According to (Sulekha, 2011), there are four main segmentation bases which includes the following; geographic segmentation (refers to geographic region, population density or climate); demographic segmentation (refers to age, sex, size and family type, etc.); psychographic segmentation (refers to life-style variables); and behavioural segmentation (refers to purchase occasion, benefits sought, user status). According to (Ogundele et al., 2013), the Nigeria marketing environment contains unique characteristics that are quite different from the developed economies and other developing countries. These include infrastructural problems, poor government support programmes for SMEs, attitude based problems on the part of the operators of SMEs and the officials of government supervisory agencies.

According to (Ogechukwu et al., 2013), The main problems of marketing encountered by SMEs in Nigeria has to do with the fact that they do not understand, and do not know how to apply marketing concept. These businesses do not have the knowledge and skills required for basic marketing. Important marketing ingredients which include marketing research, market segmentation, and marketing planning and control are often overlooked as not important. This will ultimately lead to

poor quality products, unawareness of competition, poor promotion, poor distribution, and poor pricing methods. It is therefore important for SMEs to focus on target marketing in order to grow your business. This growth involve increased sales, a larger set of employees or capital resources (land, buildings and other improvements), or the diversity and number of enterprises integrated into the operation (Thilmany, 2008). Target marketing also contributes to the design strategy for profitability and provides the chosen segments that can improve the company's competitive edge (Raicu, 1998).

The rest of this paper is structured as follows; section 2 is a review of literature, section 3 is the methodology, Section 4 is the presentation of the findings in the research, section 5 is a discussion of the findings and it also includes appropriate recommendation. Section 6 is the conclusion; Section 7 is the appendix while section 8 is the references.

2. Literature Review

Target marketing has been viewed as a major marketing concept and has been one of the main target of marketing research literature since 1960 (Wind and Bell, 2007). It is offering the right product/services to the customer at the right time, in the right location and using the proper channel. It is also a very important opportunity provided for small businesses forced to compete against larger competitors to be able to succeed. Conceptually any business strategy is usually founded on understanding, meeting and even exceeding the expectations of target segments" (Wind and Bell, 2007). The following are the components of target marketing (1) data collection, (2) application of models and frameworks and (3) resource allocation and differential action based on segment (customer) value. According to (Sulekha, 2011), the second component which is also known as Segment congruence analysis usually progresses in the following manner:

1. Fundamental dimension-reducing techniques which includes factor and cluster analysis are involved in identifying a number of segmentation bases .
2. These segmentation bases will serve as categorical variables and a multidimensional, contingency table is formed.
3. Various categorical data analysis tests will be performed on the multi-way table in order discover the nature and extent of relationships its dimensions.
4. A segmentation base is that discovered as the distinguished base and a model is developed for predicting this base from other variables.

Some approaches to discover target marketing in literature includes using Independent sample t-tests to categorizing products to understand the resulting marketing communication outcome measures (Meyers, 2005). Understanding customer needs by analysing interviews and field notes using Atlas.ti, a quantitative data analysis software application (Overton et al., 2008). Other methods of analysing data for target marketing involves testing hypotheses concerning tests of group means. These tests are performed using "parametric tests" and

include such techniques as the F, t, and z tests (Smith and Albaum, 2010).

In their research on Knowledge management and data mining, Shaw et al. 2001, recommended that, data mining will improve knowledge management process. They also observed that marketers will have better knowledge of their customers through data mining, ultimately leading to better service to customers.

According to (Radhakrishnan et al., 2013), Data mining techniques have been used to uncover hidden patterns and predict future trends and behaviors as regards target marketing in financial markets. (Rajagopal, 2011) was able to identify the high-profit, high-value and low-risk customers by using data mining technique called customer- clustering. IBM I-Miner was used to develop the algorithm. (Rossi et al, 1996) used purchase history data for target marketing using random effects choice model to measure household preferences and sensitivities. Their technique was based on Bayesian method of inference.

YongSeog (2004) proposed a data mining approach for customer targeting using artificial neural networks (ANNs) guided by genetic algorithms (GAs). Their model is able to predict an optimal target point where expected profit from direct mailing is maximized. (Abari et al., 2012) selected target market based on fuzzy analytic hierarchy process (FAHP) and technique for order preference by similarity to ideal solution (TOPSIS) methods. FAHP method was used to determine the weights of the criteria by decision target market and then rankings of the target market was determined by the TOPSIS method.

Finally, (Geloven, 2002) combined existing target selection such as linear and logistic regression, a feed forward back propagation neural network and a fuzzy modeling algorithm to analyse data for direct marketing. The strength of the approach is that the structure and specific characteristics of each feature subset are maintained and scored individually.

In 2011, Rajagopal was able to identify high-profit, high-value and low-risk customers using clustering algorithm which was done using the IBM I-Miner. In this work, further work was suggested on using more detailed behavioural data and opportunity identification using association algorithms within the segments discovered. (Raicu, 2003) proposed a theoretical data mining framework for automatic gathering of consumer data for companies and methods for discovering the relationships between customers' preferences and the product's physical characteristics. (Ayetiran and Adeyemo, 2012) was able to identifying customers who are more likely to respond to new product offers using Naïve Bayes. Finally, (Sing'oei and Wang, 2013) provided a comprehensive framework for direct marketing research using data mining and used a case study of bank marketing campaigns to evaluate the feasibility of the framework.

The following are the inadequacies of the reviewed literature; firstly, there is need to reconceptualise problems with respect to operating in the global information age and building expert system or knowledge-based systems that can help management to select and

manage the portfolio of segments (Wind and Bell, 2007), also there is need further research in market segmentation area in different specified areas to find the dominating basis of market segmentation (Sulekha, 2011) and the need to help marketers further understand how to pair up their product with visual elements in advertising that will be most conducive with meeting a marketing goal(Meyers, 2005). In addition, Rajagopal (2011), recommends further research on segmentation (clustering) using more detailed behavioral data and opportunity identification using association algorithms, (Abari et al., 2012), did not consider other multi-criteria methods to select target market. Finally, there is the need to find out how cross validation could better be integrated into the combination of linear and logistic regression, a feed forward back propagation neural network and a fuzzy modeling algorithm for directing marketing. (Geloven, 2002).

In order to attempt some of the observed research gaps, this paper aims at finding the target market of SMEs using data mining techniques. This will help in planning marketing campaigns that will make provision for effective marketing and form the basis for marketing expert systems. The problem investigated is presented in the following research questions.

1. What kind of customers is likely to be interested in doing business with a particular SME?
2. With what medium will they want to do business with this particular SME?
3. What is the opinion of the customers on the mediums investigated in no (2)?

3. Methodology

In order to carry out this research, data was collected through a survey instrument. 400 questionnaires were administered and 348 were harvested back from the respondents. The questionnaires were administered through both paper-based questionnaire and online questionnaire survey. The questions contained in the questionnaire are described in section 8 (Appendix).

In this research, the kind of customers that is likely to be interested in doing business with a particular SME is determined using sentiment analysis while k-means clustering algorithm will be used to determine the mediums that is best used to market these SMEs and the opinion of customers on the marketing mediums.

a) Sentiment analysis: The unstructured part of the questionnaire is used for the Sentiment analysis, free form textual response from the questionnaire was extracted and tokenized. The extracted tokens were then transformed into lower case which was followed by filtering of stopwords. In order to prepare the textual corpus for sentiment analysis, it was stemmed using the wordnet dictionary. All the above described process was achieved using the rapidMiner studio (https://rapidminer.com/). Sentiment analysis, also called opinion mining, is used to analyse people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions towards entities such as products, services, organizations, individuals, issues,

events, topics, and their attributes. It focuses on opinions which express or imply positive or negative sentiments (Bing, 2012). This research uses the "Extract Sentiment" operator in rapid miner studio to extracts sentiment based on SentiWordNet 3.0. The operator uses a WordNet 3.0 and a SentiWordNet 3.0.0 database to extract sentiment of an input document. The sentiment value is in range [-1.0,1.0] where -1.0 means very negative and 1.0 means very positive. (https://rapidminer.com/))

b) K-means Algorithm: Both the structured and unstructured part of the questionnaire is used to determine the best medium to market a particular SME and the opinion of respondents on the marketing mediums. According to (Satish et al., 2012), K-means algorithm is one of the most widely used hard clustering techniques.

The algorithm works as follows:

- Specify the number of clusters (k in k-means)
- Randomly pick k cluster centres in the data space
- Allocate data points to clusters according to on the shortest Euclidean distance to the cluster centres
- Re-compute discovered cluster centres by averaging the observations allocated to a cluster.
- Repeat steps three and four until convergence criterion is satisfied.

The k-means clustering was used to discover major marketing channels of selected SMEs. To achieve this, all the unstructured response is combined and rapid miner was used to cluster the data.

4. Findings

Sentiment Analysis: All the response to the unstructured part of the questionnaire was combined and sentiment analysis was carried out to generate sentiment values. Table 1 is a detailed report of the sentiment analysis result of the unstructured part of the questionnaire. This report is also presented as a line graph in Figure 1. A total of 35 SME categories were investigated and listed as follows; Auto Dealers, Gas Stations, Carwash, Tailors, Business, Consultants, Rental & Leasing, Printing &, Publishing, Communications, Marketing, Employment Agency, Educational Resources, Engineers & Surveyors, Contractors Architects, Construction, Computer Programming and Web Design, Computers & Electronics, Catering & Supplies, Movies, Event Planners, Writers, Entertainment Artists, Exercise & Fitness, Dry Cleaners, Barber & Beauty Salons,Boutique, Home & Garden, Health &, Medicine, Fast Food &Restaurant, Supermarket, Real Estate , Travels Hotel & Motel,Packaging & Shipping, Transportation. Out of these 11 were selected, as revealed in table which are the SME category that have above 50% of the investigated respondents interested in doing business with them.

From Table 1 and Figure 1, it is obvious that for all the SMEs investigated, the percentage of respondents interested in them is above 50%, this gives the platform to continue further investigation on these SMEs. Also obvious is the fact that the percentage of positive opinion for all the SMEs is low ranging between 19 to 23 percent. The table also presents the demography of the respondents that favours or not favour these SMEs

Table 1: Summary of Sentiment Analysis Results

SME Category	Total% interested in the SME at all	Total % not interested in the SME at all	%Of positive Opinion	Description of the Positive Opinion demography	%Of negative Opinion	Description of the demography of Negative Opinion
Tailors	86	14	19	Both Male and Female, Mostly Single and mainly between 21-30 years, Mostly educated with an average income	81	Both Male and Female, Both married and single, But Mostly between the ages of 31-40 and stay mostly in Lagos State
Communications	68	32	19	Mostly Male, Mostly Single, mostly above 21years with average income	81	No prominent Demography
Computers& Electronics	65	35	23	Both Male and Female, Mostly single between 21-30, average income and mostly Yoruba tribe	77	Mostly stay in Lagos
Movies	51	49	20	Mostly Male, Mostly Single	80	No prominent Demography
Exercise&Fitness	51	49	19	No prominent Demography	81	No prominent Demography
Boutique	74	26	22	Mostly B.Sc Holders with Average Income	78	No prominent Demography
Health&Medicine	75	25	19	No prominent Demography	81	No prominent Demography
Fast Food & Restaurant	71	29	18	Mostly Male, who are single between the age of 21-30	82	Both Male and Female, Between the age of 31-60
Supermarket	80	20	21	No prominent Demography	79	No prominent Demography
Travels	70	30	21	Mostly Single, Male between the ages of 21-30	79	Average Income earners
Transportation	75	25	22	No prominent Demography	78	No prominent Demography

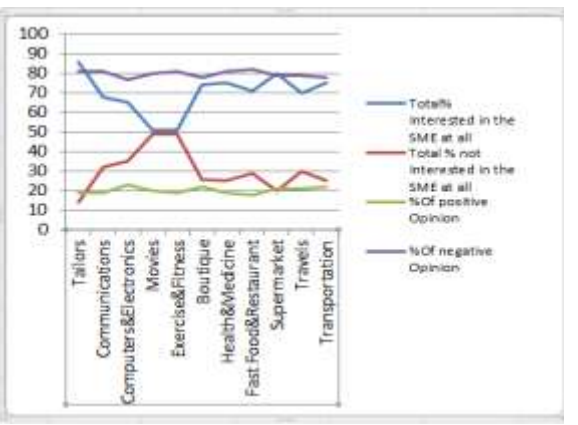

Figure 1: Graph of percentage against SMEs

Clustering: The clustering analysis result is presented as discovered from the structured data from the questionnaire response. All the attribute of the unstructured data were combined into one file and were clustered. K-means algorithm was performed 348 instances of data (total number of respondents) with 145 attributes (number of distinct words that were clustered). The algorithm ran for 10 iterations, in which 6 clusters were generated. The choice of the number of clusters is to be able to discover at least 6 major directions of the association of marketing channels to SMEs.

A summary of the six major themes talked about in the combined responses include the following as discovered from the clusters.

Cluster 0: focused mainly on the following attribute Supermarket = YesSupermarket, Barber&BeautySalons = YesBarber&Beauty, InternetSubscriptionAmount = 1000-5000, OLX = NoOLX, Whatsapp = YesWhatsApp, Yahoo = YesYahoo, Facebook = YesFacebook, EmailCheck = YesCheckEmail, Google = YesGoogle. This can be interpreted that the SMEs in the category of Supermarkets and Barger and Beauty shops will be greatly favoured if they target their clients through the following medium, Yahoo, whatsapp, Facebook, Email and google.

Cluster1: Facebook = focused mainly on YesFacebook, DoingBusinessThroughSMS = NoDoBusinessSMS, DailyCommunicationMedium = Mobile phones, Google = YesGoogle, Tailors = YesTailors. This can be interpreted as targeting SMEs in the Tailoring categories through Facebook, mobile marketing (SMS) and google.

Cluster2: focused on Google = YesGoogle, Facebook = YesFacebook, EmailCheck = YesCheckEmail, Whatsapp = YesWhatsApp, Catering&Supplies = NoCatering*number*, This can be interpreted as marketing Catering and supplies business through google, email,whatsapp and facebook.

Cluster3: focused on Google = YesGoogle, Transportation = YesTransport, Whatsapp = YesWhatsApp, Supermarket = YesSupermarket, Facebook = YesFacebook, Health&Medicine = YesHealth&Medicine. This can be interpreted as marketing Transportation, supermarket aand Health& Medicine through facebook, google and whatsapp.

Cluster4: This cluster focuses attributes such as: DoingBusinessThroug Email = NoDoBusinessEmail, Engineers&Surveyors = NoEngineers, AutoDealers = NoAutoDealers, BusinessConsultants = NoBusinessConsultant, Writers = NoEventWriters, Movies = NoMovies, Home&Garden = NoHome&Garden. This can be interpreted as people who are not interested doing business in the following SMEs…….. are also not interested in doing business through the email.

Cluster5: This cluster focuses attributes such as: Facebook = YesFacebook, Google = YesGoogle, AutoDealers = NoAutoDealers, Hotel&Motel = NoHotels. This can be interpreted as even though respondents do not do

business with AutoDealers and Hotels and Motels, they are very active on face book and Google.

The unstrucutured data used for sentiment analysis was also clustered to discover the opinions of respondents on the marketing medium. The results will also help to clarify the reason for the high negative opinion value generated in the sentiment analysis. The clustering generated 2 clusters as described below;

Cluster0: This cluster focuses attributes such as: *Single, read, dancing, particular, education, chat, cook, God, listening, family, study, idea, music, sew, play, tennis, movie, sleep*.

These are majorly hobbies, meaning these are the hobies that majority of the respondents engage in. This cluster does not reveal much as regards the reason for the high negative opinion..

Cluster0: This cluster focuses attributes such as: *fraudulence, easy, access, speed, network, failure, scam, delivery, low, wrong.*

These words reveal the reason for the high in negative opinion in the sentiment value analysis.

5. Discussions and Recommendations

In order to answer the first research question; *What kind of customers is likely to be interested in doing business with a particular SME?* We discovered that out of the 11 SMEs investigated, as presented in Table 1, young people between the ages of 21 to 30 years old are interested in doing business with the following SME categories Tailors, Computer and Electronics, Fast food and Restaurant and the Travels. These age group are not particularly opposed to any kind of business. Older people above 30 years old expressed negative opinion towards Fast food and Tailors. Also, despite the low income, average income earners are interested in doing business with Tailors, Communications, Computer Electronics and boutiques. Also of interest is that, singles males expressed high interest in Communications, Movies, Fast food and Resturants and Travels.

For the second research question; *with what medium will they want to do business with this particular SME?* We discovered that they would do business with Supermarket, Barber & Beauty Salons, through WhatsApp, Yahoo, Facebook, Email and they visit google alot. They will do business with Tailors through Facebook, SMS (mobile phones) and they visit google a lot., They will do business with Catering & Supplies through Facebook, Email, Whatsapp, and visit google a lot. They will do business with Transportation, Supermarket, Health & Medicine through facebook, google and whatsapp. They will not do business with Engineers & Surveyors, AutoDealers, Business Consultants, Home & Garden, Writers and Movies through email. Finally, Though respondents do not do business with AutoDealers and Hotels and Motels, they are very active on face book and Google.

To answer the third research question, which is, What is the opinion of the customers on the mediums investigated in research question no (2)*?* we discovered that the opinion of respondents are highly negative towards doing business with SMEs through the internet and mobile phones. Word like *fraudulence, easy, access, speed,*

network, failure, scam, delivery, low, wrong are the focus of cluster 1 in the clustering analysis, and they help to describe the reason for this high negative opinion. It is therefore clear that the reason why most respondents have reservations about these mediums is because of frauds, network failure, scam etc.

Based on the results of the analysis, it is evident that most Nigerians investigated have come on board with the advance in technology which has introduced internet and mobile ways to dong business. They find it convenient to use the internet and mobile services though they have some reservations towards this. It is therefore important that any SME who is aspiring to overcome the problems currently experienced by SMEs in Nigerian which includes the lack of understanding and application of concept needs to get on board the internet and mobile platform. This is because it is able to provide the necessary leverage required to overcome the challenges. Also these mediums will help improve the profit and competitive edge of these SMEs and thereby overcoming challenges such as poor quality products unawareness of competition, poor promotion, poor distribution, and poor pricing methods.

More specially, we make the following recommendations:

- *SMEs should have more online presence*

Presently in Nigeria today, an average SME does not have an online presence. It is therefore recommended since it is cheaper for SMEs to market themselves online, and they are also able to better target their customers, it is recommended that SMEs should be focused more on having a stronger online presence. In addition to these, they are also to create more awareness to the general Nigerian population who still have reservation getting on board the internet and mobile medium and look for ways to make their services available online at a cheaper rate.

- *Create secured transaction process.*

Apart from having an online presence, SMEs need to review their transaction modalities to be able to accommodate models which will make the customers secure enough to do business with them. Examples of these include for examples payment on delivery, verifiable address on the website. Also due to the present security concerns in the country, it is important to provide security tips to their clients to that they do not fall prey of fraudulent act for example, advise them not to meet their sales representative in an hidden and unsecured place but in a public place such and Fast food joints.

6. Conclusions and Further Study

In conclusion, the discovered result of the analysis is able to give positive direction to improving the profit of small and medium business. This is because these establishments are able to discover their target market and focus their marketing campaigns on them. Also, this study is not trivial because it will help to reduce marketing cost through target marketing. The study also helps to discover the opinion of the populace on the small and medium business and the medium in which to get to them. This discovery will definitely form solid foundation for further marketing action for them to improve their profit. For example the knowledge of the fear of online fraud can help these

businesses plan models in which to put the customers at ease if they do business with them through these medium. For further work, we plan to involve big data analysis so as to have a comprehensive source of the unstructured data effectively. According to Khan et.al., 2015, the new trends of research in data mining for target marketing is interactive data mining methods and real-time data mining.

7. References

Abari M. K., Nilchi A.N., Nasri M. and Hekmatpanah M., (2012). "Target market selection using fuzzy analytic hierarchy process (AHP) and technique for order preference by similarity to ideal solution (TOPSIS) methods" in African Journal of Business Management Vol. 6(20), pp. 6291-6299

Bing, L. (2012) 'Sentiment Analysis and Opinion Mining'. *Synthesis Lectures on Human Language Technologies.* Morgan & Claypool Publishers

CBN.,(2011). Micro Finance Policy, Regulatory and Supervisory Framework for Nigeria. A Publication of Central Bank of Nigeria, Abuja, Nigeria.

CBN (2011b). SMEs financing in Nigeria. Retrieved on the 11th of January,2011 from http://www.cenbank.org

Chukwuemeke, I. L. (2004). "Problem of Financing Small Scale Business in Nigeria". An MBA Research Publications, University of Nigeria., Nsukka.

EniolaA. A**.,** (2014). "The Role Of SME Firm Performance In Nigeria" in Arabian Journal of Business and Management Review (OMAN Chapter) Vol. 3, No.12

Gbandi E. C., and Amissah G., (2014). "Financing options for small and medium enterprises (SMEs) in Nigeria" in European Scientific Journal edition vol.10, No 1 ISSN: 1857 – 7881 (Print) e - ISSN 1857-7431

Gloven S.van., (2002). "Combining Target Selection Algorithms in Direct Marketing" http://www.tbm.tudelft.nl/fileadmin/Faculteit/TBM/Over_de_Faculteit/Afdelingen/Afdeling_Infrastructure_Systems_and_Services/Sectie_Informatie_en_Communicatie_Technologie/medewerkers/jan_van_den_berg/news/doc/sjoerd-paper.pdf

Goyat S., (2011). "The basis of market segmentation: a critical review of literature" in European Journal of Business and Management ISSN 2222-1905 (Paper) ISSN 2222-2839 (Online) Vol 3, No.9

Gulani M. G. and Usman A., (2012). "Financing Small and Medium Scale Enterprises (SMEs): A Challenge for Entrepreneurial Development in Gombe State" in Asian Journal of Business and Management Sciences ISSN: 2047-2528 Vol. 2 No. 9 [17-23]

Khan R.A., Mushtaq A. and Kanth H.(2015) Data Mining for Marketing in Journal of Marketing and Consumer Research, An International Peer-reviewed Journal Vol.9, 2015

Kim Y., Street W. N., (2004). "An intelligent system for customer targeting: a data mining approach"

in Decision Support Systems 37, 215– 228

Meyers Y. J., (2005). "Target marketing and the product: categorizing products to understand the resulting marketing communication outcome measures" in 10475 – Journal of Management and Marketing Research, page 8

Ogbadu E. E.,(2012). "Appraisal of the practical application of marketing research by SMEs in Nigeria" in Kuwait Chapter of Arabian Journal of Business and Management Review Vol. 2, No.2

Ogechukwu A. D., Oboreh J. S., Umukoro. F and Uche A. V., (2013) "Small and Medium Scale Enterprises (SMEs) in Nigeria the Marketing Interface" in Global Journal of Management & Business Research Marketing Volume 13 Issue 9 Version 1.

Ogundele O. J. K., Akingbade W. A., Saka R. O., Elegunde, A. F. and Aliu A. A, (2013)."Marketing Practice of Small and Medium Enterprises (SMEs): Perspective from a Developing Country" in Mediterranean Journal of Social Sciences MCSER Publishing, Rome-Italy Vol 4 No 3

Osotimehin K.O.,Jegede C. A., Akinlabi B. H.,and Olajide O.T., (2012). "An Evaluation of the challenges and prospects of micro and small scale enterprises development in Nigeria" in American International Journal of Contemporary Research Vol. 2 No. 4

Overton C., Volkman C., and Silver-Pacuillar H., (2008)."Understanding consumer needs through market research" in American International Journal of Contemporary Research Vol. 2 No. 4

Pusatli O. T. and Misra S.(2011) A discussion on assuring software quality in small and medium software enterprises: An empirical investigation Tehnički vjesnik Vol.18 (3), 447-452

Radhakrishnan B., Shineraj G., and Anver Muhammed K. M.,(2013). "Application of Data Mining In Marketing" in IJCSN International Journal of Computer Science and Network, Volume 2, Issue 5

Raicu, D.S. "A data mining framework for target marketing." 18th International Conference on Computers and their Applications. Int. Sc. For Computer and their Applications, Cary, NC, USA (2003) 476-9

Rajagopal S., (2011). "Customer Data Clustering Using Data Mining Technique" in International Journal of Database Management Systems (IJDMS) Vol.3, No.4

Rossi P. E., McCulloch R E., and Allenby G. M., (1996). "The Value of Purchase History in Target Marketing " in Marketing Science, Vol. 15, No. 4, (1996), pp. 321-340

Satish, G., Goutam, C., and Gary, G. (2012). 'Comparison of K-means, Normal Mixtures and Probabilistic-D Clustering for B2B Segmentation using Customers' Perceptions. *SAS Global Forum.*

Scheers L.V,(2010) "SMEs' marketing skills challenges in South Africa", African Journal of Business Management Vol. 5 No.3, pp. 5048-5056,

Shaw M.J., Subramaniam C., Tan G. W., and Welge M.E., (2001).

"Knowledge management and data mining for marketing" in Decision Support Systems 31. 127–137

Smith S. M., and Albaum G.S., (2010). "An Introduction to Marketing Research" http://cloudfront.qualtrics .com/q1/wpcontent/uploads/2012/02 / IntrotoMarketResearch.pdf

Thilmany D., "What are Niche Markets? What Advantages do They Offer?", 2015 Available: print

Wind Y. J., and Bell D. R., (2007) "Market segmentation" in The market research book, page 222

YongSeog Kim and , W.Nick Street(2004) An intelligent system for customer targeting: a data mining approach in Decision Support Systems, Volume 37, Issue 2, May 2004, Pages 215-228

8. Appendix

Structured:

1. What is your Gender?
2. What is your marital status?
3. What is your age range?
4. What is your Occupation?
5. What is your highest Academic qualification
6. What State of Nigeria do you live in?
7. What particular town in this State do you live in?
8. Are you computer literate?
9. What is your average income per month
10. What are your hobbies?
11. What is your Income range?
12. What is your tribe?
13. What is your religion?
14. How much can you afford on internet subscription per month?
15. Rate the following website according to the frequency of your visiting them. Youtube, whatsapp, facebook, twitter, google and Personalblogs.
16. Rate the following categories of Business Enterprise according to the frequency of your doing business with them.
17. How often do you check your email?
18. How often have you been convinced to do business through SMS marketing?

Unstructured:

1. What Interests you most in life?
2. What attracts you most to doing business with the select SME on the internet?
3. What do you hate most about doing business with the selected SME on the internet?
4. What attracts you most to doing business with the selected SME on the mobile phones?
5. What do you hate most about doing business with the selected SME on the mobile phone platform?

Covenant Journal of Informatics & Communication Technology Vol. 4 No. 2, Dec, 2016

An Open Access Journal, Available Online

Student Web Self-Service Portal for a Tertiary Institution

Akintola A.G.[1], Adewole K.S.[2], Mabayoje M.A.[3],& Oke J.O.[4]

[1,2,3,4]Dept. of Computer Science, Faculty of Communication & Information Sciences, University of Ilorin, P.M.B. 1515, Ilorin, Kwara-State, Nigeria.
[1]abimbolar@yahoo.com, [2]adewole.ks@gmail.com [3]mmabayoje@yahoo(gmail).com,

Abstract: The optimum condition for students to study is in an environment where they can access virtually all they need to know about courses, lecturers, locate places (for fresh and prospective students), validate staff identity, access past examination questions easily, identify vacant student hostels within and outside the school premises. Some of the reasons for students' failure can be attributed to finding accommodation within or outside the school premises. Also, fresh students find it difficult to locate specific places in the university environment and some students find it very difficult to interact with staff. This paper presents a student self-service portal to address some of these challenges. Unified Modeling Language (UML) was used to model the system. The model was implemented using Microsoft C#, Microsoft ASP.net, Microsoft SQL Server, and Google Map. The proposed system was tested and the result obtained during the execution shows that the system is capable of addressing some of the challenges confronted by students.

Keywords*:* Web Self-Service, Portal, Students, SQL, Unified Modeling Language (UML).

1. Introduction

Access to information is a serious problem faced by students (both fresh and returning students) in tertiary institutions. Each student is faced with problems to survive and succeed in higher education. However, many universities have been promoting implementation of their information systems and services to aid their students and support academic management processes in various ways.

One of the performance criteria of an educational institute is the number of successful completions by the students. However, records have shown that not all students enrolled as freshmen will succeed (Bean, 2005). The failures can

be attributed to lack of access to timely and accurate information among other factors faced by students. This is highly undesirable as this means wasted resources and reduced number of graduates to meet the demands by the industry and the community. There are many causes to this problem, and they could occur anytime during the student's course of studies. Some of these factors include; inadequate access to course outline, course notes, past examinations and test question, Lecture and Examination Time Table.

With limited resources and increasing competition for students in the education sector, higher education institutes are focusing on efforts to increase the rate of student retention (of what is being taught in the class) and completion of their academic program. Hence, the need for a web self-service portal where students can obtain information about courses, lecturers, lecture timetables etc.

Bodendorf & Sauressig (2001) defined Self Service as a way to switch from the supplier-dominated push principle to the customer-centered pull principle, in which the consumer takes an active role by initiating, controlling and tracking his/her requests.

Web Self-Service (WSS) is a type of electronic support (e-support) that allows people usually students, employees, customers etc to access information and performs routine tasks over the Internet, without requiring any interaction with a representative of an enterprise (Rouse, 2011). Web self-service is widely used in customer relationship management (CRM), employee relationship management (ERM) as well as student relationship management (SRM).

Student Web-Self Service portal is an online system that provides students with access to essential information that they will need throughout their academic careers (University of Oxford, 2013). The Web Self–Service Portal is an example of Self-Service Technologies; technological interfaces that allow customers to produce services independent of involvement of direct service employee. Self Service Technologies are replacing many face-to-face service interactions with the intention to make service transactions more accurate, convenient and faster.

2. Literature review

2.1 Definition of Self Service

It is much easier to describe self-service than to provide an exact definition of it. Self-service can be seen in stores where people use Automatic Teller Machines to withdraw, deposit or transfer funds rather than do it in banking halls, customers serving themselves in a buffet style restaurant, Petrol / Gas Stations where customers pump their own gasoline rather than have an attendant do it (this act is not popular in Nigeria), or students choosing to select and register for courses online rather than having level/academic advisers do it for them.

Many authors, journals and articles have tried to define, categorize and characterize Self Service; one of them is a particular one given by Castro, Atkinson and Ezell (2010) in their book "Embracing the Self-Service Economy", they defined Self Service as the process by which consumers engage in all or a portion of the provision of a service or product. According to Castro, Atkinson

and Ezell (2010), Self Service has long existed- think of do it yourself homeowners doing the work of professional contractors, or self-help books substituting for therapists.

Another definition given by Bodendorf & Sauressig (2001) in their book "*Internet Based Self Service Systems for Customer-oriented Processes in Public Administration*", defined Self Service as a way to switch from the supplier-dominated push principle to the customer-centered pull principle, in which the consumer takes an active role by initiating, controlling and tracking his/her requests.

Castro, Atkinson and Ezell further described how Self-Service has grown due to the many opportunities that IT has created to leverage technology for large gains in efficiency and convenience. According to them, many of these changes have become ingrained into people's way of life, for example telephone operators have been replaced by automatic telephone switching, that lets individual dial a phone number directly, elevator operators have been replaced by electronic control systems that let people operate elevators directly and at supermarkets, shoppers pick their own items rather than taking a list to a central counter and having a clerk get their goods for them.

2.2 Types of Self Service Technologies

Self-service technologies provide technological interfaces that allow consumers to use services independently of the involvement of a direct service employee. (Cardoso & Miller, 2012). Castro, Atkinson and Ezell (2010) also described self-service technology as a technology that allows a consumer to take on a role in the delivery of a service or product.

Self-service is on the rise today, because of the advances enabled by IT revolution such as the Internet and mobile phones. In addition, self-service is the natural outcome of technology that has reached maturity. When technology was new, it was often difficult to use and it required workers with specialized skills for operation (Castro, Atkinson, & Ezell, 2010). However, over-time, self-service technologies have become more user friendly to the point that the average person no longer requires a specialist to operate the technology. Castro, Atkinson & Ezell (2010) categorized channels which self-service technologies use as one of; electronic kiosks, the Internet, mobile devices, and the telephone.

2.2.1 Electronic Kiosks

Wire spring Technologies, a US based software Development Company, in their publication "An Introduction to Interactive Kiosks", described an Electronic Kiosk or Interactive kiosk as any computer-like device deployed in a public venue to give people self-service access to products and services. Kiosks are typically placed in retail stores, airports, libraries, company cafeterias and other places where personal computers are not available but self-service applications can provide some benefit. Castro, Atkinson and Ezell (2010) instead described Electronic Kiosks as devices that provide stand-alone solutions to provide users access to information or a service, such as checking an account balance at an ATM or checking in for a flight at an airport kiosk. A very important sector where

kiosks are deployed for consumer self-service is the Banking sector.

Many Banks and financial services providers offer self-service options. Automated teller machines (ATMs) are one of the earliest examples of self-service technology. First introduced in the 1970s, the technology has flourished (Castro, Atkinson, & Ezell, 2010). In a publication produced by Retail Banking Research in 2008, over 1.8 million ATMs are in operation in virtually every country, and globally consumers conduct over 44 billion transactions annually on ATMs. As technology has changed, ATMs have evolved to handle increasingly more complicated transactions and to provide customers greater convenience. Today's ATMs not only allow a bank's customers to make withdrawals, deposits, check balances, and make transfers, but ATMs may offer additional services, such as selling postage or concert tickets. Most ATMs also offer accessibility features such as voice prompts to aid visually impaired customers and multilingual options to better serve their customers (Wirespring Technologies, 2012).

Another sector where Electronic Kiosks are applied is in the Aviation Sector (Castro, Atkinson, & Ezell, 2010); Airlines have invested heavily in airport kiosks to allow customers to manage their reservations. Airport kiosks with touch screen displays, magnetic stripe card readers, and bar code scanners are now common in airports around the world. Using these kiosks, customers have the opportunity to check in for their flight, change or upgrade their seats, modify their reservation, and even purchase a ticket. Travelers without baggage can check in and then proceed directly to the gate; travelers with baggage can check in and then take their luggage to the baggage drop.

Other examples of self-service technologies that make use of Electronic Kiosks are Self-Service Gasoline Stations, Self-Pay Parking, Food-Ordering Kiosks, Airport and Travel Kiosks, Vending Machines and Reverse Vending Machines, Self-Checkout, Retail Kiosks, Human Resources Kiosks, Digital Photograph Printing, Electronic Voting, Health Care Kiosks, Information Kiosks etc. (Castro, Atkinson, & Ezell, 2010).

2.2.2 Mobile Devices

Castro, Atkinson and Ezell characterized Mobile devices as one of the most important channels for delivering self-service applications and its use will likely continue to grow in importance as wireless networks and low-cost mobile devices become more advanced. Mobile devices include smart phones, such as iPhone or Blackberry, smart cards and other portable mobile electronics.

Like Kiosks, smart phones provide another medium for interacting with online applications and services. According to Castro, Atkinson and Ezell (2010), one of the most interesting self-service applications on mobile devices is mobile commerce, a concept defined broadly as "commercial or financial transactions mediated through mobile phones or other handheld electronic devices." (Organization for Economic Co-Operation and Development, 2008). Mobile commerce is exploding worldwide with research firm Juniper predicting that, by 2011, the global value of all commercial or financial

transactions effected through mobile phones will exceed $587 billion. Much of this is driven by browser-enabled smart phones that allow individuals access to any Internet-based application from a mobile device. In addition, many self-service applications currently available online have been, or will be, adapted for smart phones, for example, Banks have developed mobile applications so that customers can access their funds, transfer money out of their accounts, pay for bills etc. from their mobile phones (Lomas, 2008).

According to Investopedia an online investment encyclopedia, another important self-service technology is mobile payments systems; this involves using a mobile phone as an electronic wallet; a system that securely stores user's payment information and passwords for numerous payment methods and websites. Combined with near field communication (NFC) technology, a specific standard of Radio frequency Identification (RFID) technology, NFC-capable phones can securely transmit data wirelessly over short ranges between electronic devices thus enabling contactless payments (Castro, Atkinson, & Ezell, 2010).

2.2.3 Telephone Applications

IT also enables consumers to use the telephone to access self-service solutions. In particular, telephone operators have been largely replaced with digital technology. The major reason why productivity for telephone operators has increased approximately 12 percent a year since 1950 is because customers, rather than operators, now place the vast majority of calls through direct dialing (Castro, Atkinson, & Ezell, 2010). In addition, when requesting a phone listing, most consumers use a technology that allows the phone company's computer to ask the customer to say the listing they want, saving an operator from asking that. Voice recognition technology is getting so effective that there is little need for the operator to be the go-between for the customer and the telephone company computer

2.3 Applications of Web Self Service

Indeed, various Internet applications have equipped consumers to take on new roles and responsibilities that previously required assistance from individuals employed in the service sector, including professionals from virtually every field from banking to education to retail. Internet based Self Service has been applied in various sectors, According to Castro, Atkinson, & Ezell, they include; Online Health, Online Banking, E-Learning, Professional Services, Retail E-Commerce, Customer Service, Ticketing and Reservations etc.

2.3.1 Online Health

In HealthCare, IT empowers patients by giving them access to the latest medical research, their own health records, and information on the quality of care they receive. Online applications such as Microsoft Health Vault have emerged to allow individuals to track and analyze their personal health information. The Microsoft HealthVault is a web-based platform from Microsoft to store and maintain health and fitness information. (Rouse, 2011). The HealthVault allows users to upload data from a small number of HealthVault-compatible devices, it also allows sending, receiving and storing of medical records

and information from doctors and healthcare providers. Also, there are online applications that allow patients to track health markers such as their blood pressure, cholesterol, and body mass index to see how these indicators change over time and how they compare to healthy patients of the same age and sex. As a result, patients are less dependent on health care workers for medical solutions and can take more active role in their own care. Examples of these applications are Revolution Health, WebMD etc. (Castro, Atkinson, & Ezell 2010).

2.3.2 Online Banking

Another area in which Web based self-service technologies has manifested is also in the banking sector, much like ATMs, online banking has replaced the need for tellers for most financial transactions at a bank. Banks increasingly offer online applications that meet the day-to-day banking needs of most customers, such as opening an account, checking account balance, and transferring funds. Self- Service technology has become so matured that Nigerian banks like Guaranty Trust Bank, United Bank for Africa now provide facilities for opening bank account on their respective Facebook Pages. Other banks such as WEMA bank, Stanbic IBTC have followed suit by providing reliable internet banking facilities that allow their customers to transfer funds, pay bills etc. on their websites.

2.3.3 Retail E-Commerce

In the area of E-Commerce, Self-Service technology gives consumer control over their service encounter (Castro, Atkinson, & Ezell 2010). Buying goods and services online allows consumers the freedom to choose where and when to shop, and the opportunity to research the product, the seller and any other available options. Just about anything that can be bought in a store can be bought online, even perishables like groceries. And consumers have embraced these possibilities; more than 85 percent of the world's online population has purchased something using the Internet. (The Nielsen Company, 2008). According to Castro, Atkinson & Ezell, the Internet has introduced many online services that substitute physical goods for digital goods. Online services like iTunes, Amazon, Jumia, Konga and the E-bay allow consumers to find and purchase goods without ever interacting with a service worker.

2.3.4 Customer Service

Many companies provide self-service options for customers to receive customer service online. The service options range from a simple list of frequently asked questions to advanced online applications that give customers access to detailed information and services so that many of them can solve their own problems. Instead of consumers contacting customer service representatives, they can go online and do the work themselves and find a solution in the same time or less. According to Castro, Atkinson and Ezell (2010), companies like FedEx and UPS allow customers to track their packages online rather than call a customer service agent to find out its status, Also Computer manufacturers like Dell allow customers to look up product information and get support based on the unique serial numbers printed on each device. Many businesses also give

their customers online access to their accounts. For example, utility companies, cable companies, and telephone service providers all typically offer online access so that their customers can pay their bills online, see past statements, and make changes to their service. Some companies have gone a step further with online self-service and created human-like automated customer service agents (Castro, Atkinson, & Ezell, 2010).

2.4 Benefits of Self Service

Based on the already explained types of self-service technologies, Castro, Atkinson and Ezell (2010) noted that self-service has grown so largely, as it provides benefits to consumers, organizations, and the economy as a whole.

Self-service technology can provide consumers greater convenience, accessibility and ease of use. According to a paper published by Oracle Corporation in 2013 "Benefits of Web Self-Service", the company noted that consumers want to solve their issues quickly and easily on the Web, when they can, they are likely to buy more, with 80% saying they will increase their spending.

According to Castro, Atkinson and Ezell, Convenience is a big factor; Self-service technologies often make business available 24 hours a day, seven days a week rather than being limited to traditional working hours. They also noted that, consumers often find self-service technology empowering; using the technology, the consumer can control the service encounter and not feel rushed or pressured.

Self-service technology can also make service encounters more accessible for individuals with disabilities. For example, individuals with mobility disabilities may find online shopping more accessible than shopping in brick and mortar (Castro, Atkinson, & Ezell, 2010). Individuals can take advantage of accessibility options in Web browsers to access online applications and services.

For businesses, Castro, Atkinson and Ezell (2010) noted that investment in self-service reduces cost and helps provide a better quality product or service. According to them, using self-service technology frees up workers that can either be reassigned to more profitable jobs or eliminated to reduce pay-roll costs. For example, a retailer that introduces self-checkout can reassign cashiers to sales or customer service jobs to increase sales and customer satisfaction or cut these jobs to save on overhead.

The economy also benefits from self-service technology, According to Castro, Atkinson and Ezell (2010); Per-capita income growth is the single most important indicator of a nation's economic well-being. And per-capita income is largely a function of the growth of productivity (the amount of output per hour of work). Higher productivity growth goes a long way in solving pressing societal problems, including lagging income growth, national debt, and the ability of society to spend in key areas (e.g. transportation, environmental protection, and health coverage).

2.5 Concerns Over Self Service

In spite of the significant benefits of self-service, particularly for economic growth, self-service has some criticisms or concerns. Castro, Atkinson and Ezell (2010), identified four major concerns, they are listed below;

- Self-service simply shifts work to the consumer with only the company benefiting.
- Self-service eliminates consumer choice and robs individuals of human contacts.
- Self-service eliminates jobs and finally,
- Benefits of self-service will not go to workers.

The following concerns on implementing a self-service system are also noted;

- Errors: The rate of errors is likely to increase.
- Training and Support: Users of a self-service system must have training before they can use the technology.
- Cost: Implementing a self-service system requires an upfront investment.
- Security Concerns

2.6 Student Web Self-Service Portal

Portals started as a way to organize a variety of web-based information sources on one desktop interface: a search tool, news feeds, links to favourite web sites, content organized by topics, and so on.

Portals in tertiary institutions serve multiple functions for multiple customers with one tool. Lee J., Hong, N. L. & Ling, N.L. (2002) noted that a portal provides the faculty members at the individual campuses with efficient, direct links to current knowledge about teaching and learning through technology among the campuses of the university system, nationally, and internationally.

The trend toward portals as the technology tool of choice for knowledge sharing leads to the convergence of knowledge management. As Kidwell, Linde and Johnson (2000) stated, portals have the potential to revolutionize learning in tertiary institutions and have ways of encouraging: -

i. Increased competitiveness and responsiveness for research grants, contracts, and commercial opportunities.
ii. Reduced turnaround time for research.
iii. Minimized devotion of research resources to administrative tasks.
iv. Facilitation of interdisciplinary research.
v. Leveraging of previous research and proposal efforts.
vi. Improved internal and external services and effectiveness.
vii Reduced administrative costs.

However, some of the limitations of portals is the fact that they are costly and require a lot of time for maintenance and management (Olibie and Akudolu, 2015).

Narrowing it than, very good examples of a student web self-service portal are university portals. The portal allows students to perform variety of actions, some of which include;

i. Registration of Courses.
ii. Payment of charges.
iii. Viewing and Printing of Results.
iv. Editing of Personal Details.
v. Access to Current School Information etc.

However, needs of students of the Universities are not fully met by these portals, below are some of the limitations of the existing portals:

i. Locating Specific Places in the University Campuses is a headache especially for fresh students as there is no diagrammatic representation of the Campus Layout on the school portal for students to navigate through the campus.
ii. The current portal takes care of hostels available on campus which is limited in size and cannot contain the ever increasing student's populace.

3. Methodology

A Software Development Life Cycle (SDLC) is essentially a series of steps or phases that provide a model for the development and lifecycle management of an application or piece of software (Glynn, 2008). The Software Development Model used in carrying out this development is the Agile Development Model. The processes under this model are:

a. Communication (Requirement Gathering and Specification)
b. Design and Development
c. Integration and Testing
d. Operation and Maintenance

3.1 Communication

This is the first stage of the software development process; it includes requirement gathering and requirement specification. The interview style of gathering information was adopted.

After interacting with several students, it is discovered that there is need to develop a web portal that will be able to perform the following services:

a. Allow students to locate specific places on Campus.
b. Allow students access to past questions and course materials.
c. Allow students access to available students accommodation on and off campus
d. Allow students access to Staff's profile via a search app, that will enable students view a staff's profile by basically typing the staff's name in a textbox that will be provided on a page in the portal.

3.2 Design and Development: This phase includes the class diagram, user flow diagram, administrative flow, and the entity relationship diagram.

3.2.1 Class Diagram: The diagram shows the link between the department and faculty and relation to the user. It also provide for staff of the department and courses offered along with past question for the courses.

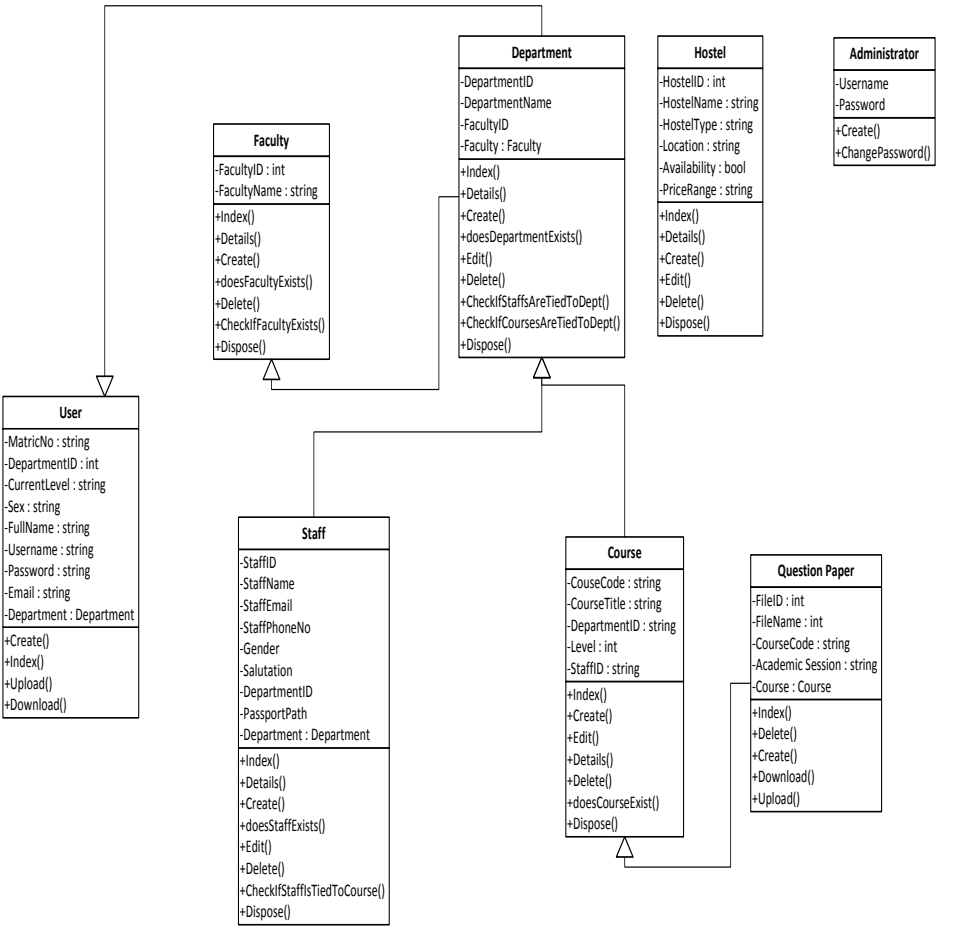

Figure 1. Class Diagram

3.2.2 User flow Diagram: This is the user flow chart diagram. In between the start and end, the user can perform several operations like register, log in , download and so on.

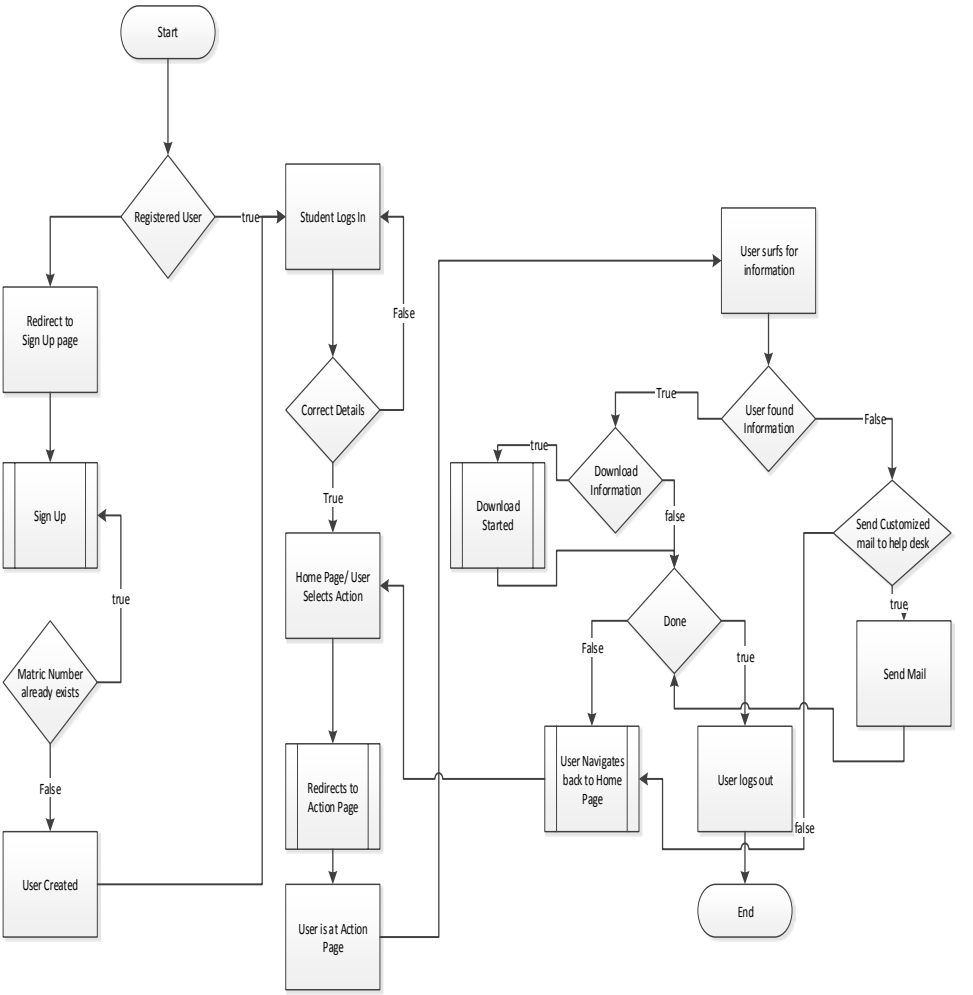

Figure 2. User Flow Diagram

3.2.3 Administrator Flow: The administrative flow chart log in page from which other actions can be performed.

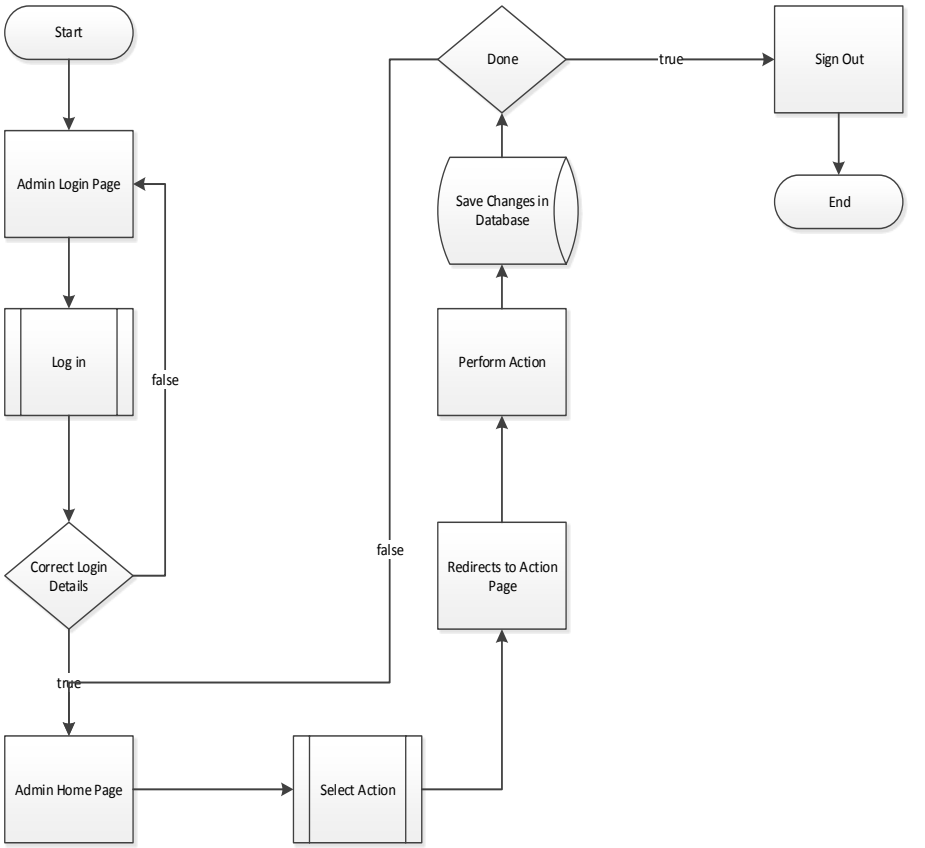

Figure 3. Administrator flow

3.3 Entity Relationship Diagram: The diagram shows the user page with the user having access to department, faculty, courses, staff etc.

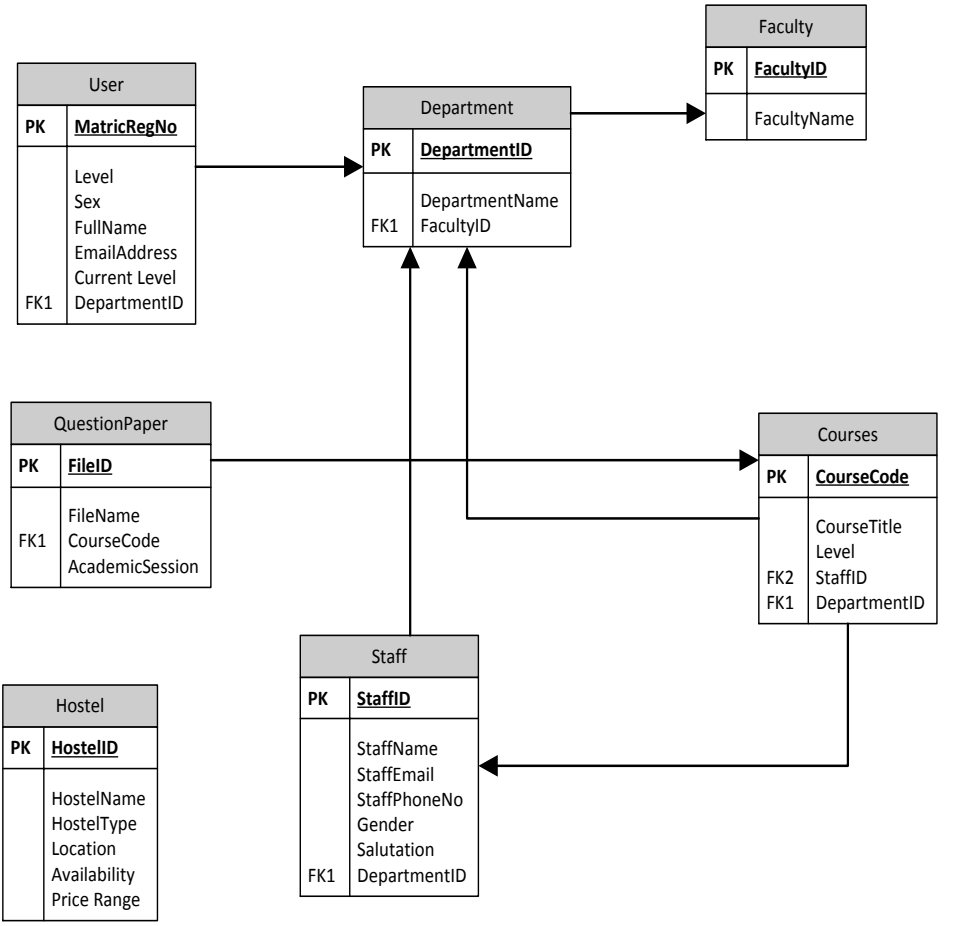

Figure 4. Entity Relationship Diagram

4. Implementation and Results

User Home Page: This is the first page the user of the application sees, it features an interactive Map of the University and links to performing different functions on the portal.

University and links to performing different functions on the portal.

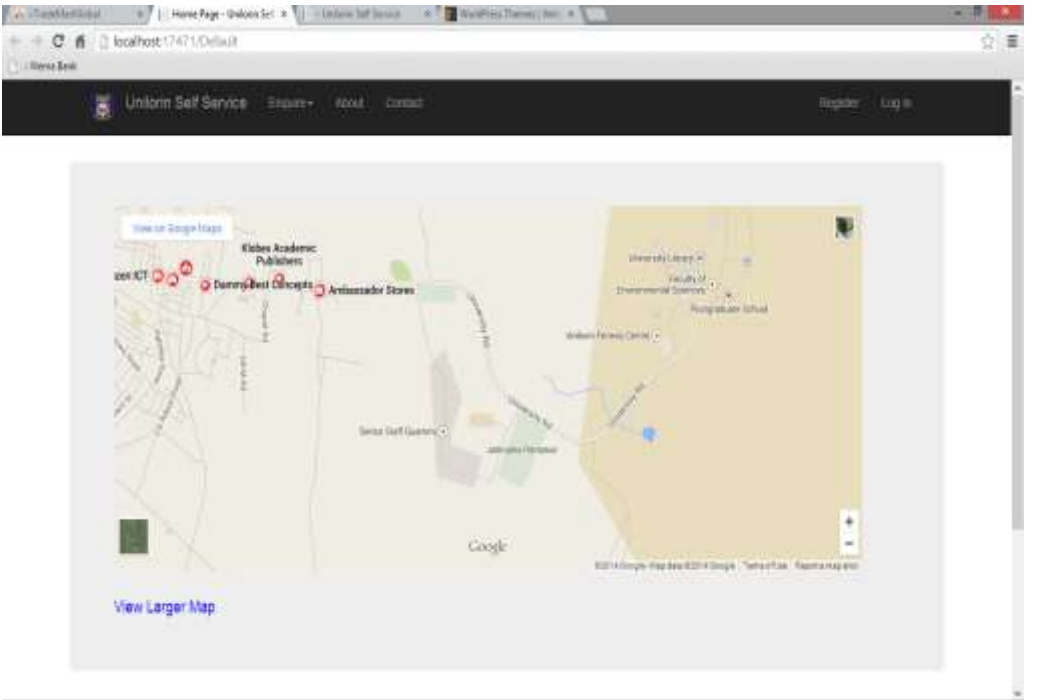

Figure 5. User home page

Portal Registration Page: This is the page where the user of the portal registers his/her details for portal authentication. Details collected on this page are the User's Matric Number or Registration Number, the User's current level, Sex, Faculty, Department, Choice Password and Email Address.

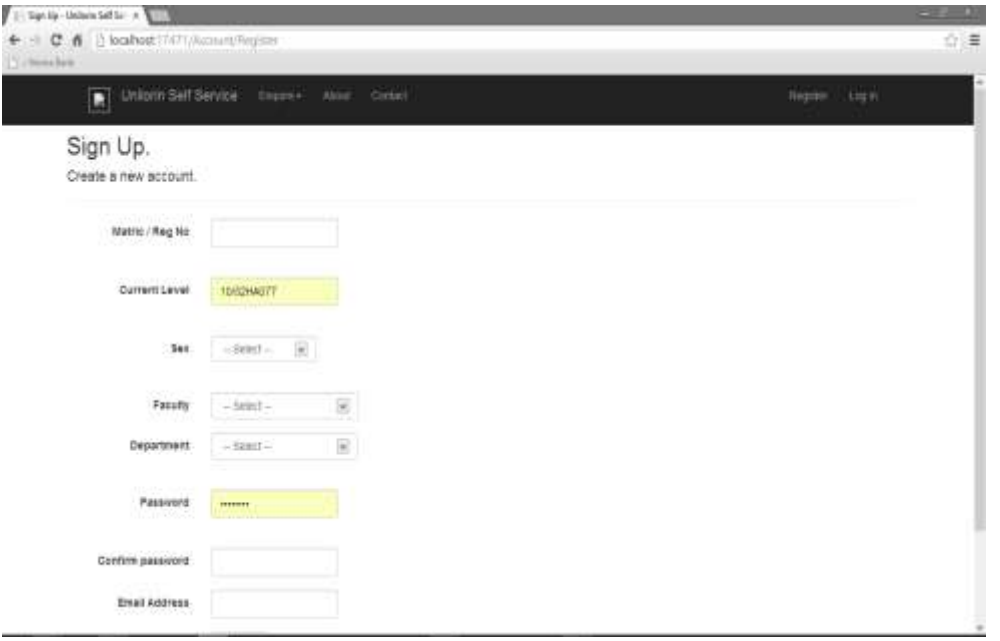

Figure 6. Registration page

User Login Page: This is where the User Login to the portal in order to access functionalities of the portal.

The details required to login are a registered email address or matriculation number and password

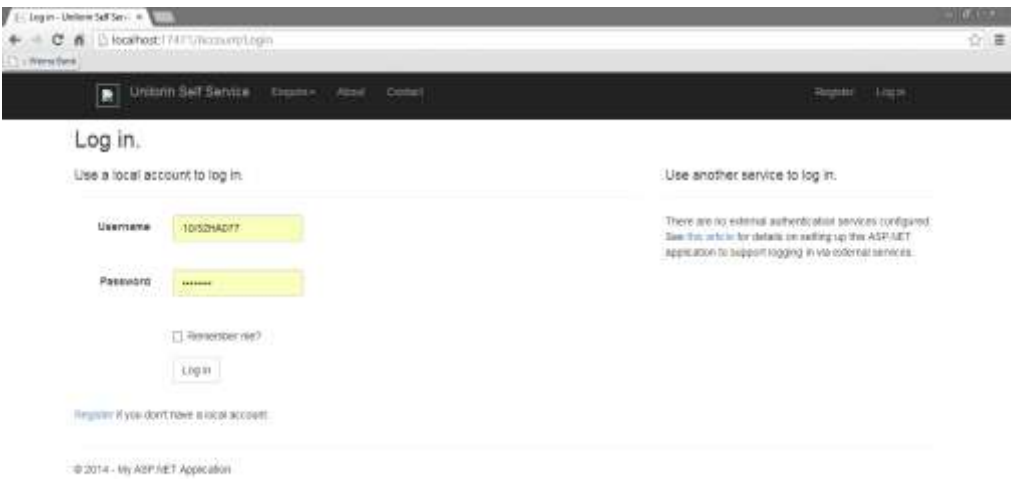

Figure 7 User login page

Staff Search Page: These allow the student to search for any member of staff from different faculties.

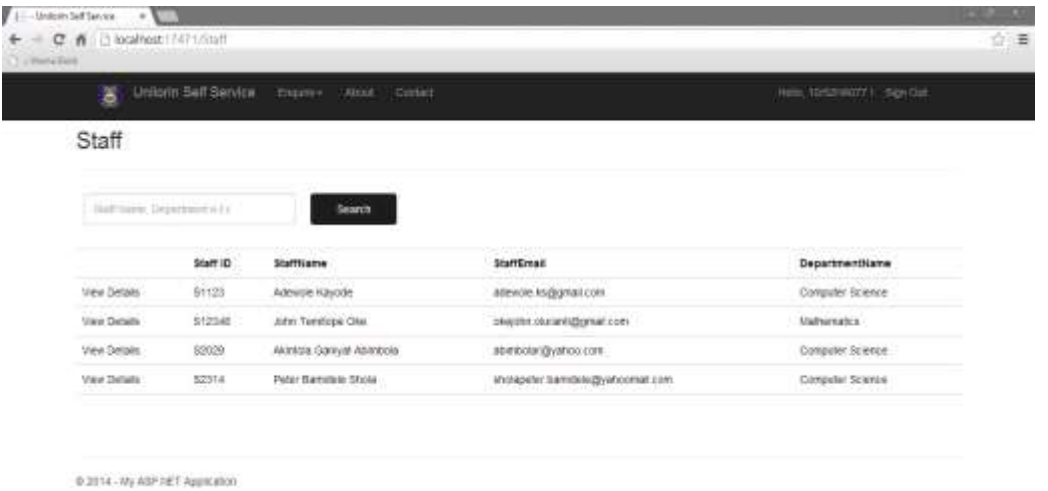

Figure 8. Staff search page

Staff Details Page: Page showing details of a particular staff including His/her Staff ID, Salutation,

Name, Email Address, Phone Number, Gender, Department and Faculty.

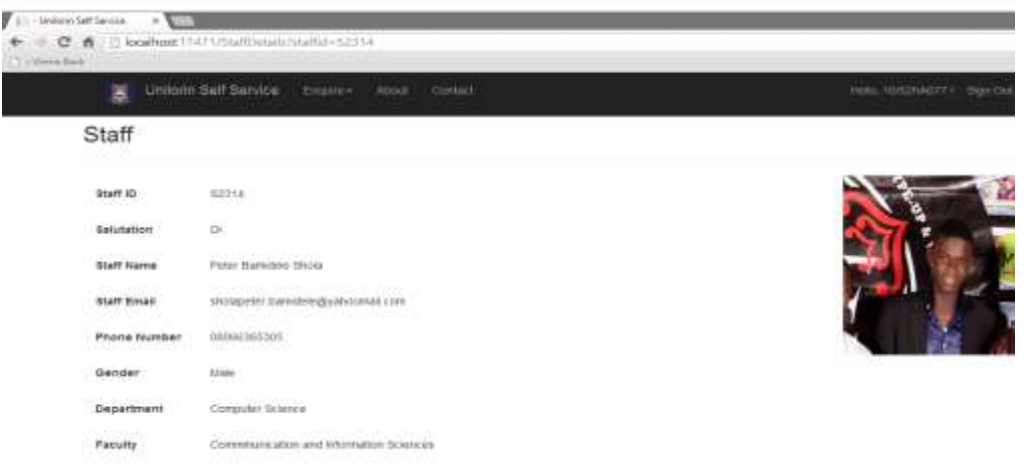

Figure 9. Staff details page

Course Search Page: On this page, the course enters the course code or course title he/she wants to access and the query fetches the details from the database.

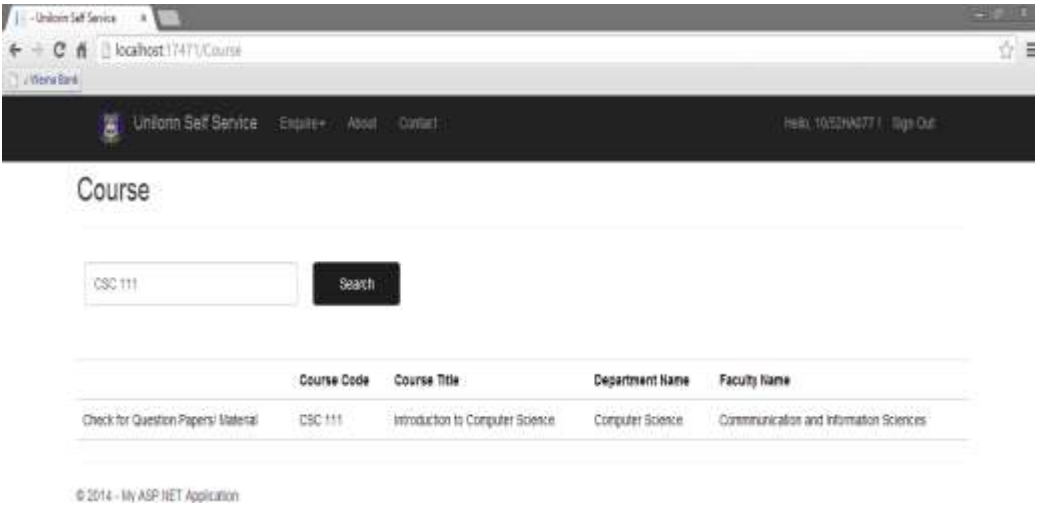

Figure 10: Course Search Page

Course Details Page: Page showing details of a particular course including available materials for download.

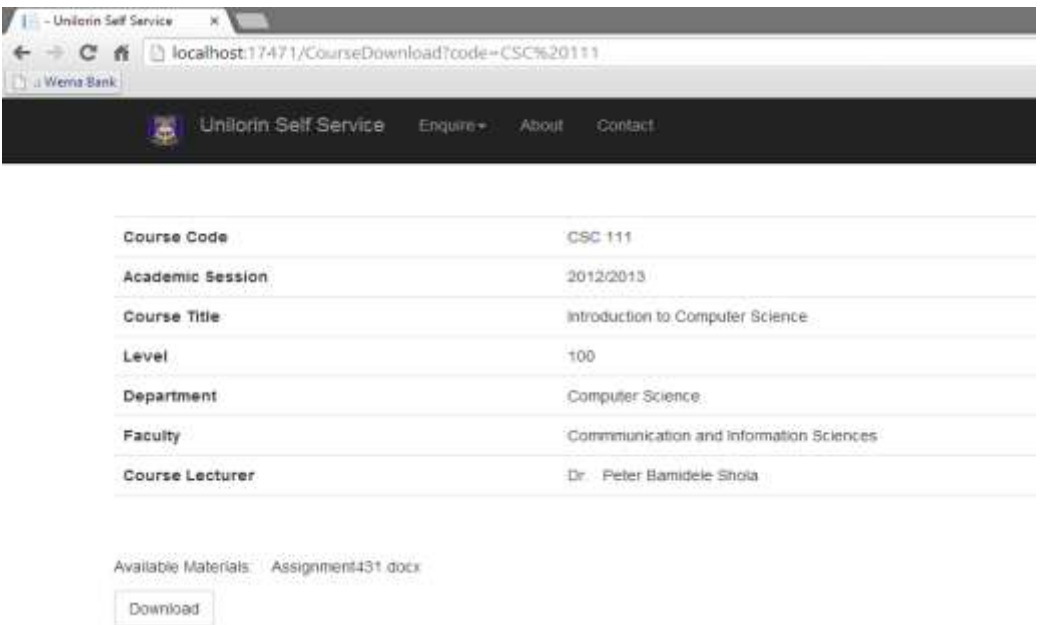

Figure 11. Course search page

5. Discussion

The work ascertains and describes how Online, Web or Internet-Based Self-Service can be applied to Students in a University Environment. The study rigorously visited various scholarship orientations and relevant literature as well as research on the concept of Self-Service and its applications.

The developed web portal model will be able to perform the following services over existing ones in such a way that it allows students to: locate specific places on Campus, access past questions and course materials, access available students accommodation (both) on and off campus and access Staff's profile via a search app, that will enable students view a staff's profile by basically typing the staff's name in a search tool that will be provided on a page in the portal.

6. Conclusion

A student web portal has been developed using the concept of Self-Service. In achieving this, an existing portal; the University of Ilorin Undergraduate portal was evaluated and current limitations were identified. The Project work ascertains and describes how Online, Web or Internet-Based Self-Service can be applied to Students in a University Environment.

An Agile Development methodology is chosen for developing the portal because agile methodologies emphasizes on working software and changing requirement over robust documentation or analysis. Technologies used in implementing the portal include the Unified Modeling Language (UML), Microsoft C#, Microsoft Visual Studio, Model-View-Controller Architecture, ASP.net, JQuery, LINQ, SQL, Microsoft SQL Server, Google Map etc.

7. Recommendations

The current implementation can be extended and improved in many ways. The portal can be extended to meet the needs of other entities in the University for instance, Staffs as the portal is designed mainly for the usage of students of the University.

Also, the features present in this portal can be integrated into the already existing academic portal of the University.

References

Bean J. P. (2005). Nine Themes Of College Student Retention. *College Student Retention: Formula for student success*, A. Seidman, Ed. America: CT:Praeger, 2005, p. 29.

Bodendorf and Sauressig G., (2001). *Internet Based Self Service Systems for Customer-oriented Processes in Public Administration.* Nuremberg, Germany:CSC PLOEZENKE AG.

Cardoso J. and Miller J. A., "Internet-Based Self-Services: from Analysis and Design to Deployment," Proceedings of the 1st IEEE International Conference on Services Economics (SE'2012), in conjunction with the 2012 IEEE International Conference on Web Services (ICWS'12), Honolulu, Hawaii (June 2012) pp. 24-31.

Castro, D. , Atkinson, R., & Ezell, S. (2010, April). Embracing the self-service economy. Retrieved from:http://www.itif.org /files/2015-self-service.pdf.Computing

Careers. (2006). *Software engineering*. Retrieved from www.computingcareers.acm.org: http://computingcareers.acm.org/?page_id=12

Glynn, F. (2008). *Software Development Life Cycle*. Retrieved January 5, 2013, from www.veracode.com: http://www.veracode.com/security/software-development-lifecycle

Group 3. (2013). *CSC 316: Software Engineering Project Write Up.* Ilorin, Nigeria.

Hsin, T. (2000). Online shopping: A Fad or A Revolution. In *Retail World Asia 2000* (p. 25). Retail Group Malaysia.

ISTQB GUID. (n.d.). *What are the Software Development Models?* Retrieved January 5, 2014, from www.istqbexamcertification.com: http://istqbexamcertification.com/what-are-the-software-development-models/

Jarvenpaa, S., & Todd, P. (1997). *Consumer Reactions to Electronic Shopping on the World Wide Web*. International Journal of Electronic Commerce.

Jawahar. (2010, December 8). *Overview of System Analysis & Design.* Retrieved January 3, 2014, from www.ddegjust.ac.in: http://www.ddegjust.ac.in/studymaterial/pgdca/ms-04.pdf

Kelley, E. J. (2009). *The Importance of Convenience in Consumer Purchasing.* Retrieved from JSTOR: http://www.jstor.org/discover/10.2307/1248014?sid=21105570593991&uid=2&uid=4&uid=3738720

Kidwell, R.P., Linde, M. & Johnson, T. (2000). *A survey into the application of ICT for educational purposes in higher education* (pp. 176 – 181). Rottrtdsm: Erasmus Universiteit/OECR.

King, & Juhn-ShiuanLiou, S. F. (2004). *A Framework for Internet Channel Evaluation.* International Journal of Information & management.

Koehler, J., Nebel, B., Hoffman, J., & Dimopoulos, Y. (1997). *Extending Planning Graphs to an ADL subset. Lecture Notes in Computer Science.*

Lee, J., Hong, N.L. & Ling, N.L. (2002). An analysis of students' preparation for the virtual learning environment. *The Internet and Higher Education, 4, 231 – 242.*

Lomar, T. (2003). Fast content switching with respect 2.0 Ericsson gmbh.

Makafan, P. (2013). *CSC 316: Software Engineering Class Project Group 3 Write Up.* Ilorin, Nigeria.

Margherio, L. (1998). Secretariat for Electronic Commerce. In *The Emerging Digital Economy*. Washington:: US Department of Commerce.

Microsoft Developer Network. (n.d.). *Visual C#*. Retrieved Janaury 5, 2014, from msdn.microsoft.com: http://msdn.microsoft.com/en-us/library/kx37x362.aspx.

(OECD), O. O. (1997). *Dismantling the Barriers to Global Electronic Commerce.* Paris: OECD.

Olibie, E. and Akudolu, L. (2015), Paper accepted for publication in East African Journal of Educational Research and Policy (EAJERP).

Oracle Corporation (2013). Benefits of web self-service. *Right now cloud service*.

Plan, P. G. (2011, November 8). *Difference Between Shopping Cart and ECommerce*. Retrieved from prettygoodplan.com: http://prettygoodplan.com/2011/11/

difference-between-shopping-cart-ecommerce/

Rouse M. (2011); Web self-service definition. Retrieved 23/8/15: http://searchcrm.techtarget.com/definition/Web-self-service.

Rouse, J. (2011). Articulating the World: Experimental Systems and Conceptual Understanding *International Studies in the Philosophy of Science* 25 (3):243 - 254

Rowley, J. (2000). Product Search in E-shopping: A Review and Research Propositions. In *Journal of Consumer Marketing* (pp. 124-135).

Tatum, M. (2013, November 25). *What is a Web Portal?* Retrieved December 15, 2013, from www.wisegeek.org: http://www.wisegeek.com/what-is-a-web-portal.htm

University of Oxford. (2013, November 13). *Student Self Service*. Retrieved December 12, 2013, from www.ox.ac.uk: https://www.studentsystem.ox.ac.uk/

Wikipedia. (2011, November 16). *Self Service Technologies*. Retrieved December 12, 2013, from www.wikipedia.org:http://www.wikipedia.org

Covenant Journal of Informatics & Communication Technology Vol. 4 No. 2, Dec, 2016

An Open Access Journal, Available Online

A Hybrid Web Caching Design Model for Internet-Content Delivery

Alfa A. A[1], Ogwueleka F. N[2], Dogo E. M[3], Sanjay M[4]

[1&3] Federal University of Technology, Minna, Nigeria
[2] Federal University Wukari, Nigeria
[4]Covenant University, Nigeria
Contact(s). [1]abrahamsalfa@gmail.com , [2]drfrancisca@yahoo.com, [3]eustace.dogo@futminna.edu.ng, [4]ssopam@gmail.com

Abstract - The need for online contents (or resources) to be shared and distributed in a large and sophisticated networks of users, geographical dispersed location of servers and their clients, time taken to fulfil clients requests pose major challenge. Therefore the choice of suitable architecture for Internet-based content delivery (ICD) technologies readily comes to mind. To achieve this, Akamai and Gnutella Web technologies are extensively reviewed to identify their strengths and weakness because of their popularity across the world for delivering contents. This new design for Internet-based content distribution is called AkaGnu because of the extra layer (Gnutella network) inserted into Akamai architecture, which provides greater Internet edge over each technology deployed independently. The paper presents a new ICD technology that performs better than Akamai system as a result of new features and behaviours introduced that reduce network traffic, more clients Internet connectivity, increase file sharing, improved speed of contents deliveries, and enhanced network security.

Keywords/Index Terms- ICD, Akamai, Gnutella, peer-to-peer, AkaGnu, network traffic, security, architecture, technology

1. Introduction

The Internet supports different kinds of services, such as content delivery (David, 1998). The Internet is world-wide network (or collection) of thousands of computers and computer networks. This idea was first conceived by an American Scientist Vinton Cerf as a collaborative enterprise with the United State Department of Defence Advanced Research Projects Agency (ARPA) in 1973 supervised by Robert Kahn (Ion *et al*, 2003).

Traditionally, systems for delivering content have been designed to support the client-server architecture; a case is the World Wide Web (WWW). But, recent development on the Internet content delivery field has changed due to greater awareness of a new application such as peer-to-peer (P2P) file allocation/sharing. Systems with P2P are fully distributed, utility for proprietary naming, encoding mechanism and protocols (Ion *et al, 2003;* Antony and Peter, 2001; Sylvia *et al,* 2001; Ben *et al,* 2004).

Internets are made up of local networks connection with the help of special computers in each local network (referred to as gateways) (Vinton, 1973). Gateways interconnections are established through various communication channels such as telephone lines, optical fibres, radio and satellite links for the purpose of information exchange and delivery to remote machine using computerized address of that specific machine. The Internet protocol (IP) is the underlying software that is used to control the Internet that is, it determines how information are routed using the gateway machines from the computer sending information to the computer receiving, while Transmission Control Protocol (TCP) ensures the information sent has reached the destination computer and, otherwise, it causes the retransmission of the information (Vinton, 1973).

Content delivery (sometimes referred to as content sharing or caching) is a service of reproducing website page(s) to servers that are physically isolated. Also, Internet content is composed of bits and data packets that are shared or distributed in the Internet. It includes a piece of news or a hypertext make-up language (HTML) page, video or sound files (Zhang, 2010). This is mostly deployed by high-traffic web site manager owners and Internet Service Providers (ISPs) to improve content delivery especially when request for a page is made, it identifies dynamically and serves the page content available on the server closest to the requester (or user) (Vinton, 1973; Zhang, 2010).

The techniques for efficient content delivery include: deployment of cache servers with considerable size access points to Internet located throughout the world; and the use of specially designated routing code that sends request for a web page to the nearby server. Others include redirection of HTTP, Internet Protocol (IP) and domain name system (DNS) forwarding (Zhang, 2010).

The use of content delivery has been effective for specialized events with high-traffic such as Live Web broadcast that involves content to be continually shared or dispensed from the originating server through satellite links (or connectivity) to the receiver server (Vinton, 1973; Zhang, 2010). To that end, there has been an increased attention for enhancing mechanisms used to transport Internet content from servers-to-clients. One of the basic mechanisms currently deployed to improve the distribution of static Internet content is called proxy caching. Multimedia content streaming involves proxy caching and multicast delivery. These operations to a larger extent reduce network load (Zhang, 2010).

Several models exist for a provider seeking to distribute content. The client server is the most deployed in basic Internet. Others are Cloud and Content Delivery Network (CDN) that offer comprehensive packages of services to content providers apart from content distribution. More so, P2P model and other several architectures of information networking are instances of Internet-based content delivery (Zhang, 2010).

Content delivery is now made easy with the use of the Internet because it is fast, convenient and reliable. The technologies of content delivery using the Internet, its benefits and challenges as well as ways of improving content delivery were discussed in this research work. A portion of Afribank's Content delivery network was used to study the behaviour of workloads and traffic.

2. Related Works

Previous Research works on Internet and Content delivery focus on Web caching performance. There are three main features that make workload characteristics interconnected to Web caching, which is the most widely researched aspect of workloads. First, algorithms design for both Web proxy caching and cooperative Web. Proxy caching has been an enormously vigorous area of research in recent years (Chankhuntod *et al* , 1996*;* Zhang *et al*, 1997; Touch, 1998; Fan *et al*, 1998; Karger *et al*, 1999; Tewari *et al*, 1999). Second reason that makes it a topic of interest may be because of usefulness of caching as a technique for performance improvement in other parts of computer systems such as file system and architecture of central processing unit (CPU). Third, Content Distribution systems and Web caching are two most commonly deployed applications presently to improve performance of the Web (Tewari *et al*, 1999).

Operationally for proxy caching to be successful, Alastair (2002) proposes that the workload must have the following attributes: the caches allow storage for documents with repeated access requests, those documents that do not alter in between the repeated accesses and many of the requested document have been previously requested at some point in the past (Alastair, 2002). However, we have found that three workload attributes are not available for any kind of CDNs (such as Akamai and Gnutella), and most essential to the overall success of Web caching that is the cachability of Web documents, the rate of changes to Web document, and the amount of reuse of Web document (Alastair, 2002).

The basic Internet largely supports one computing model, the client-server model (Lewandowski, 1998). In the client-server model a server (or a pool of servers) stores information and services and waits inactively for the clients to make request for them. The client-server model needs components of network to function effectively. These network components are placed between the clients and the server for structured communication (that is, structure of the basic Internet interconnectivity). The servers and clients connect to the Internet access providers, who in turn connect to the Internet backbone provider for connection to the whole network. These together with the software employed in

the network make up the infrastructure of basic Internet, which can be exploited for Content delivery (Alastair, 2002).

Another widely used Internet Content delivery model is the cloud computing. Cloud computing is a service for management of hardware and software in a better and easier manner. Clouds are made up of pools of virtualized resources such as hardware, software and services that can be accessed with ease (Vaquero *et al*, 2009). The motive behind the cloud is to relocate the network infrastructure, which reduces the cost of resources management for better scalability and flexibility of network. The three major services rendered by cloud are: Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) (Zhang, 2010).

Content Delivery Network (CDN) provides generally all-inclusive Services, when it is compared to cloud computing. The CDN was intended as overlays, which are virtual topologies that are positioned on top of the basic Internet. It enhances the value to the Internet for instance, guaranteeing the retrieval of data as well as balancing of load (Doval and O'Mahony, 2003). Some of the values introduced by a CDN are reliable network, minimized latencies for consumers, improved throughput and balancing of origin server load (Vakali and Pallis, 2003). This means that, the end-to-end connection is abolished, but, two distinct end-to-end connections exist though, one between the client and the CDN, and the other between the CDN and the Content server. The services offered by CDNs are full storage and distribution as well as management of software and hardware for simplifying Content delivery (Zhang, 2010).

Schollmeier (2002) provides another kind of basic Internet known as peer-to-peer network, which consist of dispersed resources linked by network. The major advancement over the client-server model is that clients connected to the network can perform the role of servers, which is not the case in the client-server because, the role of clients and servers are discrete.

Few things can be said to match with the growth of Internet over the years, especially its phenomenal growth perhaps in the past ten years. The main problem faced by Internet has been infrastructure for distributing progressively more intricate data to a larger and increasing users' population. The idea to further scale down is the motivation for the design of thousands of clusters of node, global-scale CDNs and lately, independent P2P structures. These techniques of Content Delivery have swiftly altered the composition of Internet Content Delivery (ICD) and traffic; hence appreciating the present-day Internet requires for a thorough appreciation of these latest mechanisms as well as the kind of data they serve (Sefan *et al*, 2002).

3. Content Delivery Technologies

There are three main ways of categorizing Internet content/workload distribution namely; the World Wide Web, a content delivery network (Akamai), and peer-to-peer systems (Gnutella).

3.1 The World Wide Web

Tim Berners-Lee in 1989 put forward that, World Wide Web is a content delivery network with client-server

architecture, in which a centralized server contains all the content and accessed through the client's browser (Sefan *et al*, 2002). All the clients in the network are identified using Internet Protocol (IP) addresses. All the contents are saved on Web Server using unique Web addresses as content identifiers. Web workloads include objects with a variety of different types such as text, images, archives, executable code, audio, and video data. Though, the makeup of content types are largely objects (such as text and images), while the remaining content types are relatively small portion of resources. There are two implications of the predominance of text and image data in Web workloads. Firstly, text is amenable to compression. Secondly, text and image data are highly cacheable types of content. A large fraction of Web traffic is therefore likely to be cacheable. The Hypertext Markup Language (HTML) is the standard representation for hypertext documents in American Standard Code for Information Interchange (ASCII) format. HTML allows publishers of content to format their content, reference images or other objects, and embed hypertext links to other content (Sefan *et al*, 2002).

Generally, most of these transmitted objects are small in size, while some other objects are very large. Many workloads exhibit a common variance because, most of the objects are small in size accounting for a relatively small portion of the bandwidth consumed. But, a small quantity of usually very large objects is the justification for a considerable size of the bandwidth consumption. Recently, with emergence of new applications which have a sudden and profound influence on the distribution of content types. There is increasing availability of multimedia content in Web workloads. A higher portion of Web workloads is dedicated to transporting multimedia data. By implication, the size of Web objects could change over time. A different trend that can affect Web workloads is the penetration of high-speed Internet connections (or broadband). Users with high-bandwidth Internet connections are likely to download larger resources (Sefan *et al*, 2002).

The Web, Akamai, and Gnutella are being currently deployed systems architecture for content distribution on the Internet. The main goal of our system when compared to these systems is very similar, that is to deliver content on the Internet. Of course, these systems have well-elicited notions of clients, servers, and objects. These systems follow the same basic principle: clients fetch objects from servers, but with varying architectures. Figure 1 shows the conventional architecture of a content delivery network in which a single server is responsible for delivering content to many users across the Internet. Each arrow indicates a client's query for an object located on the server (Sefan *et al*, 2002; Geoff, 1999; John *et al*, 2002; Jem, 2003).

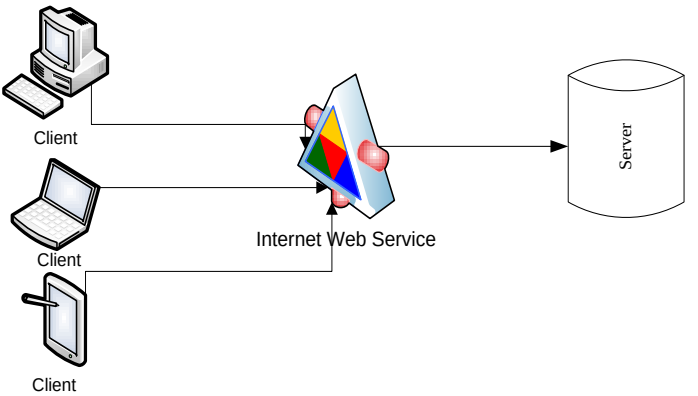

Figure 1: **The conventional layout of client-server architecture of Internet content delivery.** *Source*- (Zhang, 2010)

AkaGnu network consists of a naming infrastructure, a document representation language and an RPC protocol. It uses Uniform Resource Locators (URLs) for naming content. For instance, if a URL is http://postgraduate.futminna.edu/download.htm. This can be broken down in to three parts as shown in Table 1.

Table 1: A URL Parts Resolution

URLs PARTS	MEANING
http	The protocol for communicating with server
Postgraduate.futminna.edu	Server's name
Download.htm	The name of a file on that server

3.2 Content Delivery Network

Content delivery network is made up of a group of non-originating servers that make effort to download work from origin servers by distributing content in place of origin server (Sefan et al, 2002). The servers belonging to a CDN are typically placed at diverse locations throughout the network, whereas some or all of the origin server's content are cached (or replicated) among the CDN servers. For each request, the CDN makes effort to trace a CDN server nearest by proxy to the client to fulfil (or carryout) the request, where the impression of close could be described as latency, geographical or topological.

Comparing CDN to systems based on the client-server or the P2P architecture, a limited number of content delivery networks exist on the Internet, high traffic due to direct inter-communication between its clients and the server. These networks are suitable to be run and managed by private companies. Akamai infrastructure is a typical example of CDN (Sefan et al, 2002; Geoff, 1999; John et al, 2002; Jem, 2003).

3.2.1 Akamai Technology

Akamai technology is a business-related content delivery network consisting of hundreds of thousands of content distribution servers dispersed throughout the world whose roles are to distribute content to nearby clients. Web servers sign up with Akamai to replicate and serve a portion of their local content from the Akamai's servers. The benefits from using Akamai are two-fold. First, Akamai servers perform as proxy caches. They help to minimize client latencies, network traffic and server load. Second, proxy caching service rendered by Akamai is accessible to all clients on Internet world-wide, as opposed to proxy caches that only serve a limited client population (Sefan et al, 2002).

Akamai exploits DNS-based name redirection to forward (or route) requests of client to Akamai servers. Upon receipt of a DNS request, an Akamai nameserver provides the corresponding address of an Akamai content server located closest to the client issuing the request. Because of the transparent nature of DNS name resolution, Akamai's client redirection mechanism does not require any modifications to client software, server protocols, or Web applications. Akamai's architecture: Web client's requests are routed to nearby Akamai servers instead of the origin Web server as illustrated in Figure 2. Web is similar to Akamai, because it delivers HTML objects, it uses URLs to name objects, and it uses HTTP to transport content (Stefan *et al*, 2002).

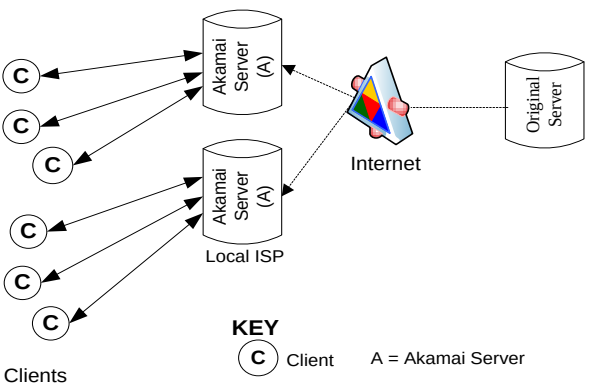

Figure 2: **Akamai network architecture.** *Source*-(Alastair, 2002)

3.2.2 Gnutella Peer-to-Peer Network Technology

Gnutella is a dispersed peer-to-peer system, made up of hosts linked to one another over TCP/IP and running on common software that supports the Gnutella protocols. This connection of individual hosts (or nodes) produce a network of computers exchanging Gnutella traffic such as queries, replies to queries as well as other control messages used to detect nodes. Basically, this network permits the participating hosts to easily share

arbitrary resources. For instance, some resources may be associated with other resources, meta-information (or other sorts of pointers). At present, its exploitations are based exclusively on data files; that is a specific host can submit its local files for others to download, and can download files put forward by others on their computers for the purpose of sharing. The control of each node is the responsibility of user running the application software, who partakes in the network by: (a) specifying explicitly a list of local files to share across the network (b) searching for files existing somewhere on the network and (c) downloading files from other nodes (Matei, 2001; Sefan *et al*, 2002; Stephen, 2004).

Gnutella is a network of broadcast-type that pings and queries are duplicated and relayed to several other nodes. To minimize consumption of network resource, nodes cache pongs and supply them as responses to pings when they can. Pongs and Query Hits are routed by each node back along the path needed to reach the destination. Another important attribute is Ultra peer scheme that increases efficiency and scalability of network by classifying nodes into normal clients and super nodes. A super node is a reliably connected host with sufficient network bandwidth that performs as a proxy for a well-built numbers of connecting clients. The super node eliminates the trouble of extensive network message routing from client that may be a low bandwidth modem user. In this case, the modem user (leaf node) makes use of the well-connected super node as network's access point.

Gnutella network imitate the Internet itself; nodes of low bandwidth are connected to well-built super nodes (routers) that help pass on majority of the data over high bandwidth backbones (Matei, 2001; Sefan *et al*, 2002; Stephen, 2004).

Gnutella architecture is shown in figure 3.

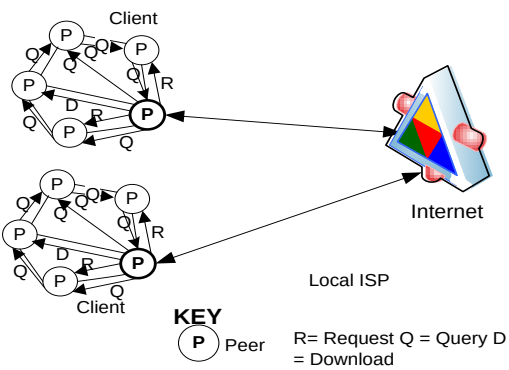

Figure 3: **Gnutella peer-to-peer network architecture.** *Source-* (Zhang, 2010)

4. Methodology

The new model is developed based on limitations observed from Akamai's infrastructures such as latencies, traffic congestions and file sharing. Using Akamai's infrastructure, each client will have to query and get response directly from Akamai server (or local ISP) which leads to lower network performance. AkaGnu network design is composed of Akamai's infrastructure and Gnutella peer-to-peer infrastructure to enhance the network performance for content delivery as shown in Figure 4.

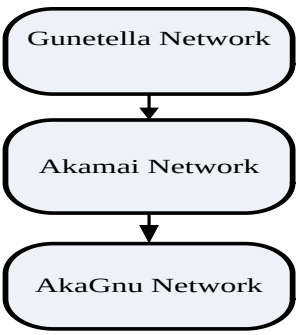

Figure 4: Design steps of AkaGnu Network

All peers exhibit client-server characteristics, to find a file; a controlled flood of a query packet is initiated by a peer across the overlay network to all of its neighbours (or peers). A peer checks upon taking delivery of a query packet, if the query matches any locally stored files. If confirmed, the peer sends back a query response packet towards query originator peer. In a case where a file match is not found, the peer sends continuously a huge number of the query across the overlay. HTTP protocol is used to download content, once content is found.

AkaGnu architecture is made up of two system architectures, that is, peer-to-peer client (Gnutella) to the right hand and the content delivery network (Akamai) as shown in Figure 5.

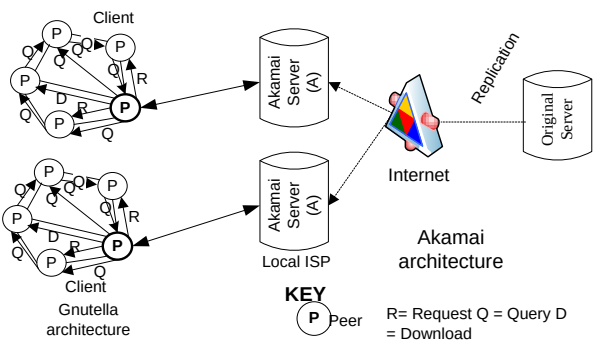

Figure 5: AkaGnu technology architecture.

The original server replicates (or copies) its contents through the Internet platform to Akamai servers (meta-data (a set of data about an entity)/information) located in the Local ISP. On the other hand, the clients are connected to decentralized peer-to-peer network with a host running Gnutella software providing protocols. All the peer-to-peer clients have access privileges to query (Q) and respond (R) to queries and upload/download (D) data files. These processes are controlled by a super node (or client) connected to both clients peer-to-peer and the Akamai servers in the Local ISP. The new system is different from other systems because of improved security, reduced network traffic on the original server and Akamai servers, file sharing among all peer-to-peer clients without internet connections, minimized rate of clients accessing network resources and fast and increased availability of contents.

5. Results and Discussions

5.1 Monitoring Traffic Generation of AkaGnu Vs Akamai and Gnutella Networks

AkaGnu uses Gnutella protocol versions 0.4 (from the legacy peers), the ultra-peer and leaf peers are implemented. The function of the protocol is to provide bootstrapping and simulator runs application as a component of the AkaGnu technology. The legacy server afribank.local provides the Akamai server the contents of the original server for the bank's branches transactions processing and the resources. The two branches of the bank were connected each to a local Akamai network (a sub-server) of the original server located at the head office. Each branch has a leaf node and super node system typical of Gnutella architecture. Every branch uses IP addresses to initiate communication to every other node and super node such as messages passing, query and respond, download and upload, file sharing and resource sharing without having to connect directly to the main Akamai servers. The peer-to-peer nodes communicate with each other directly and Akamai server through the super nodes, as result the traffic congestion experienced by branch staff connected to these branches local network is reduced, support for larger volume of transactions and need for continuous local and Internet connections.

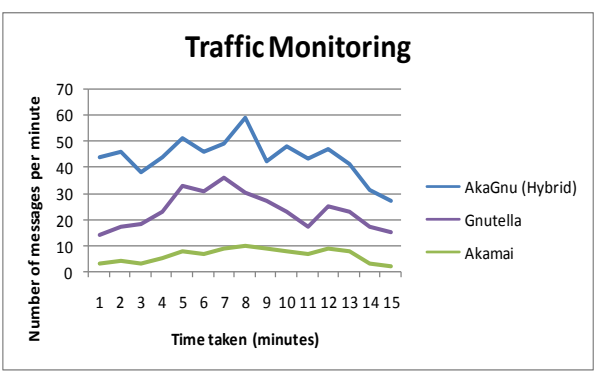

Figure 6: **Traffic monitoring for Akamai, Gnutella P2P and AkaGnu (Hybrid) Networks.**

Figure 6 shows the mean number of messages and time taken for 20 trials for Akamai, Gnutella P2P and Hybrid CDNs monitored over 5 months period, December, 2012 to April, 2013, at Afribank's ICD networks. The mean numbers of messages are plotted against time taken to deliver the messages over the CDNs separately for Akamai, Gnutella P2P and Hybrid networks, to enable us differentiate the effectiveness of the each of the CDNs.

It can be observed that there is a clear difference in the number of messages delivered at a particular time for each of the ICD techniques, with Akamai recording the smallest number of messages and requiring longer time for successful delivery as illustrated in Figure 6. Gnutella P2P is better when compared to Akamai in terms of time required to deliver messages over its network, because peer-to-peer infrastructure provides faster Internet and content delivery. The Hybrid (AkaGnu) is the best ICD technology, because it allows more messages delivery within a relatively short time as shown in Figure 6. The results of hybrid technology for ICD far outweigh the individual technique of Akamai and Gnutella P2P networks, because peer-to-peer architecture, relationships existing among clients, speed of Internet and local ISPs connectivity, reduced traffic congestions and availability of contents on the network.

5.2 Analysis of Node Connectivity and Network Topology

One major attribute of the AkaGnu network over 4 month period compared to Gnutella network (as illustrated in Figure 5) is the scaling down of the network magnitude. The average number of connections for every node is higher and relatively stable for AkaGnu network when compared to Gnutella network connectivity as shown in Figure 7. This behaviour makes it feasible to determine the number of connection a larger network will generate and finding limits of scalability as based on bandwidth available. When evaluating connectivity and reliability patterns globally in AkaGnu network, it is essential to point out the self-organized behaviour of the network. Users decide on only the maximum number of connections a node should maintain, while nodes based on local information only decide to whom to connect or when to drop/add a connection.

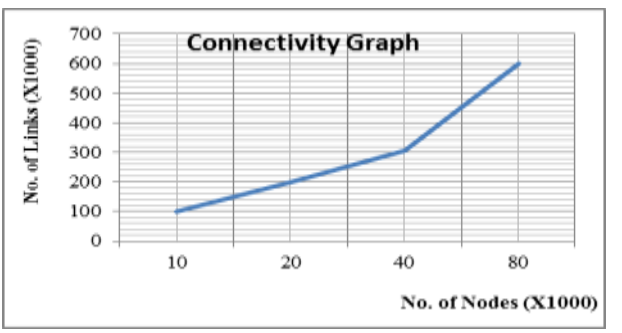

Figure 7: AkaGnu (Hybrid) Node connectivity relationship.

5.3 Benefits of a Hybrid CDN Technology

Some of the benefits of the hybrid CDN include:

a) It uses http as its file-transfer protocol; it implements a micro and standard Web browsers which facilitate easy access to other peers and their files.
b) It provides end-users improved performance such as in data availability, reduction of server load and load balancing.
c) It has further scaled down the workloads associated with the Akamai CDN by decreasing the amount of time taken for requests (or queries) to get responses accomplished by Akamai servers.
d) It allows sharing of files among peers of clients directly, minimizing storages required for floods of queries, contents and responses.

6. Conclusion

AkaGnu is a web caching Internet-based technique for content delivery. AkaGnu technology has further scaled down the workloads associated with the Akamai content delivery network by reducing the amount of time taken for requests (or queries) to get responses accomplished by Akamai servers, allow sharing of files among peers of clients directly, minimizing storages required for floods queries, contents and responses. There is a lower hit rate on the Akamai servers and originating server. Traffic on the Internet and local ISP is significantly minimized as a consequence of shorter path travelled by queries/responses in the networks. The security of the Hybrid design model is better, because it combines best security features of Gnutella and Akamai networks. This design model guarantees better performance, availability of contents and services any time desired without delays. The limitation of AkaGnu network is the cost incurred in building each Gnutella P2P that is attached to the Akamai network in the Hybrid architecture. The security of files and individual system on the Hybrid CDNs is another future research area.

7. References

Antony R. and Peter D. (2001). Pastry: Scalable, Distributed Object Location and Routing for Large-Scale Peer-to-Peer Systems. In proceedings of IFIP/ACM Conference on distributed systems platforms (Middleware). Heidelberg, Germany. 294(4). Pp. 329-350.

Alastair W. (2002). Sharing and Caching Characteristics of Internet Content. Doctor of Philosophy Thesis of University of Washington, USA. Pp. 1-170. Retrieved from http://www.google.com.ng/content-delivery.htm

Ben Y. Z, Ling H, Jeremy S, Scan C. R, Anthony D. J, and John K. (2004). Tapestry: A resilient global-scale overlay for service deployment. IEEE Journal on Selected Areas in communications, 22(1): 41-53.

Chankhuntod A, Danzig P, Neerdaels C, Scwartz M and Worrell K. (1996). A Hierarchial Internet Objects Cache. In Proceedings of the 3rd USENIX Technical Conference. San Diego, CA. Vol. 3. Pp. 153-164. Retrieved from URL

http://www.rp.lip6.fr/ieeecomms97.pdf

David D. C. (1998). The Design Philosophy of the DARPA Internet Protocols. In Proceedings of the ACM SIGCOMM Technical Conference, Stanford, CA. Vol. 6, No. 2. Pp. 117-129.

Doval D and O'Mahony D. (2003). Overlay Networks: A Scalable Alternative for P2P. IEEE Internet Computing, Trinity College Dublin. 7(4), Pp. 79-82.

Fan L, Cao P, Almeida J and Broder A. Z. (1998). Summary Cache: a Scalable Wide-Area Web-cache Sharing Protocol. Technical Report 1361, Computer Science Department, University of Wisconsin. Pp. 1-17. Retrieved from http://www.google.com/web-cashing.htm

Geoff H. (1999). Web Caching. The Internet Protocol Journal. Cissco Systems Inc. Publications and Machandise. CA, USA. Volume 2, No. 3. Pp. 2-19.

Ion S, Robert M, David L, David K. M, Frans K. K, Frans D. and Haris B. (2003). Chord: A scalable peer-to-peer lookup protocol for Internet applications. IEEE/ACM Transactions on Networking 11(1): 17-32.

Jem E. B. (2003). Decentralized Peer-to-Peer Network Architecture: Gnutella and Freenet. A Technical Paper of University of Manitoba, Winnipeg, Manitoba. Canada. Pp. 1-9.

John D, Bruce M, Jay P, Harald P, Ramesh S. and Bill W. (2002). Globally Distributed Content Delivery. IEEE Internet Computing. Akamai Technologies. Pp. 50-58. Retrieved from http://www.akamai.com/d/technical publications/GloballyDistributedContent Delivery.pdf

Karger D, Sherman A, Berkheimer A, Bogstad B, Dhanidina R, Iwamoto K, Kim B, Matkins L, Yerushalmi Y. (1999). Web Caching with Consistent Hashing. In Proceedings of the 8th International World Web Wide Conference, Elsevier North-Holland Inc. New York, USA. May 1999. Vol. 31. Pp. 11-16.

Lewandowski S. M. (1998). Frameworks for Component-Based Client/Server Computing. ACM Computing Surveys, 30(1), Pp. 157-175.

Matei R. (2001). Peer-to-Peer Architecture Case Study: Gnutella Network. Computer Science Department, The University of Chicago, USA, Technical Report ITR 0086044. Pp. 1-7.

Schollmeier R. (2002). A Definition of Peer-to-Peer Networking for the Classification of Peer-to-Peer Architectures and Applications. In: Graham R. L., Shahmehni N, Proceedings of the first International Conference on Peer-to-Peer Computing (P2P 2001), Linkoping, Sweden, Vol. 1. Pp 27-29.

Sefan S, Krishna P. G, Richard J. D, Steven D. G. and Henry M. L. (2002). An Analysis of Internet Content Delivery Systems. University of Washington Thesis submitted to Department of Computer Science and Engineering. USA. Pp.1, Retrieved from http://www.google.com.ng/content-delivery.htm

Stephen B. (2004). Monitoring the Gnutella Network. In proceedings on Intelligent Agents, University of Technology, Sidney, Australia, Pp. 1-10.

Sylvia R, Paul F, Mark H, Richard K, and Scott S. (2001). A Scalable Content-addressable Network. In proceedings of the ACM SIGCOMM 2001 Technical Conference, San Diego, CA. USA. Vol. 12. Pp. 161-172.

Tewari R, Dahlin M, Vin H, and Kay J. (1999). Design Considerations of Distributed Caching on the Internet. In Proceedings of the 19th International Conference on Distributed Computing Systems (ICDCS), IBM Faculty Award, Austin, Texas, USA. May 1999. Vol. 45, Pp. 273-284.

Touch J. (1998). The LSAM Proxy Cache- A Multicast Distributed Virtual Cache. In Proceedings of the 3rd International WWW Caching Workshop, Manchester, England. Pp. 92-108. Retrieved from URL http://eprints.kfupm.edu.sa/70445/1/70445.pdf

Vakali A and Pallis G. (2003). Content Delivery Networks: Status and Trends. IEEE Internet Computing, Society Journal. Arnetminer, Greece. 7(6), Pp. 68-74.

Vaquero L. M, R. L, Caceres J. and Lindner M. (2009). A Break in the Clouds: Towards a Cloud Definition. ACM SIGCOMM Computer Communication Review, 39(1), Pp. 50-55.

Vinton C. (1973). Fascinating facts about the Invention of the Internet. Pp. 1-3. Retrieved from http://www.ideafinder.com/history/inventions/internet.htm

Zhang L, Floyd S and Jacobson V. (1997). Adaptive Web Caching. In a proceedings of the Second International Web Caching Workshop. Atlanta, GA. Pp. 1-9. Retrieved from URL http://www-ngr.ee.lbl.gov/web/html

Zhang N. (2010). Internet Content Delivery as a Two-Sided market. Master's Thesis of AALTO University. Pp. 7-16. Retrieved from http://www.google.com.ng/content-delivery.htm.

Covenant Journal of Informatics & Communication Technology Vol. 4 No. 2, Dec, 2016

An Open Access Journal, Available Online

Monitoring Cardiovascular Disease-Patients with Mobile Computing Technologies

Zacchaeus Oni Omogbadegun[1] & Adesewa Taiwo Adegoke[2]

[1,2]Computer and Information Sciences Department, College of Science and Technology, Covenant University, Ota, Ogun State, Nigeria

[1]zacchaeus.omogbadegun@covenantuniversity.edu.ng [2]dgkadesewa@yahoo.com

Abstract: Physicians and healthcare networks have been slow to adopt electronic medical records and to integrate medical data with the ubiquitous mobile device. Mobile and wearable systems for continuous health monitoring constitute a key technology in helping the transition of health care to a more proactive and affordable healthcare. Cardiovascular Disease (CVD) includes dysfunctional conditions of the heart, arteries, and veins that supply oxygen to vital life-sustaining areas/organs of the body. CVD singly accounts for about 40% of all deaths worldwide. Over 80 per cent of CVD deaths take place in low- and middle-income countries. An estimated 17.5 million people died from cardiovascular disease in 2005, and expected to top 20 million per year by 2015. By 2030, more than 23 million people will die annually from CVDs. CVDs' patients face risks of recurrent acute cardiovascular events, hospital re-admission, and unfavourable quality of life. Heart Failure, (HF), leads to death if not properly managed and supervised. Current treatments for Congestive Heart Failure (CHF) provide a limited palliative outcome. New technologies are now pertinent to generate high-dimensional data that provide unprecedented opportunities for unbiased identification of biomarkers that can be used to optimize pre-operative planning, with the goal of avoiding costly post-operative complications and prolonged hospitalization. Due to the crucial role of remote monitoring for CVD patients, significant efforts from research communities and industry to propose and design a variety of CVD monitoring devices have become imperative. This paper builds a proof-of-concept and presents a cardiovascular monitoring system, Cardiovascular Disease Management System (CVDMS), for real-time information on patient's heart health status with respect to his/her heart beat in hemodynamics computation towards reducing re-admission incidence problem. Administered 485 questionnaires and interviewed 12 cardiologists, 45 physicians, and 23

pharmacists to gather details on vital CVD parameters. 469 of 485 questionnaires (96.70%) were validly completed and returned, while 16 (3.30%) were not. Searched internet databases and cognate texts for literature. A mobile CVDMS for HF was developed using UML, MySQL Server 5.0, Java servlets, Apache Tomcat 6.0 server, microcontroller, and Ozeki sms server. Patient completes a questionnaire on a J2ME platform-based computing device that measures the heartbeat rate. Biological signals acquired by CVDMS are processed by microcontroller. Pulses are counted within a space of one minute to know heartbeat rate per minute. The CVDMS application gets the heartbeat reading, and if the heart rate is abnormal, a trigger is set enabling the Ozeki SMS Gateway to send an alert to patient's next-of-kin and cardiologist. CVDMS guarantees individual patient's direct involvement to closely monitor changes in his/her vital signs and provide feedback to maintain an optimal health status. Medical personnel get alerted when life-threatening changes occur in establishing proper communication between patient and cardiologist via sms. Hemodynamics computation could be performed with the parameters obtained from the data supplied by CVDMS as a cardiovascular intervention to save many lives and improve quality of life.

Keywords: artery stiffness; blood pressure; cardiologist; cardiovascular disease; heart attack; heart failure; hemodynamic volumetric parameters; hospital re-admission; hypertension; risk-factor.

1. Introduction

The cardiovascular system consists of heart, vessels, and blood. In a healthy person, the heart pumps the blood in vessels with synchronous pulses (HR) and pulse wave velocity (PWV). The source of power of life is the heart, and the blood nourishing the body constantly flows under her impetus. However, she also demands the nourishing of blood. Coronary artery, namely three blood vessels respectively located in the heart, can supply blood and oxygen to her. The coronary artery is the artery special for supplying blood to the heart. If cholesterol and other substances are accumulated in the blood vessels, the vascular cavity will be narrower or be blocked and the blood flow will be smooth and then be blocked to cause cardiac ischemia and a series of symptoms which are coronary heart disease, namely coronary atherosclerosis. Coronary heart disease (CHD) is also called as coronary atherosclerotic heart disease. The excessive fat deposition results in atherosclerosis and weakened elasticity. The mortality of human on cardiovascular and cerebrovascular diseases induced on the arterial vessel wall has exceeded 1 / 2 of the total mortality of population. Dangerous factors making the elasticity of coronary artery weakened are high blood fat, smoking, diabetes, obesity, high blood pressure, lack of physical activity, psychological overstrain, family history of coronary heart disease, oral contraceptive, etc. The force of blood flux, which is caused by heart beating, forms a pressure against blood vessels' walls. Blood pressure, (BP), is a vital measurement used by the physicians for diagnosing the health situation of subjects, and saving them from critical diseases or some dangerous circumstances, such as hypertension,

hypotension, artery stiffness, coma or heart attack (Al-Jaafreh and Al-Jumaily, 2008). Cardiovascular Disease (CVD) includes dysfunctional conditions of the heart, arteries, and veins that supply oxygen to vital life-sustaining areas of the body like the brain, the heart itself, and other vital organs. Cardiovascular disease, including heart disease and stroke, remains the leading cause of death around the world. CVD being the prime cause of death among the elderly in industrialized countries is a major determinant of chronic disability. Cardiovascular diseases represent the main cause of death for people of developed countries, and frequently they may account for premature fatal outcomes even in the apparently healthy young. The morbid entities are mostly structural, affecting the major components of the heart (aorta, pulmonary artery, pericardium, coronary arteries, myocardium, endocardium and conduction system). Yet, most heart attacks and strokes could be prevented if it were possible to provide an easy and reliable method of monitoring and diagnostics. In particular, the early detection of abnormalities in the function of the heart, called arrhythmias, could be valuable for clinicians (Nataraj et al. 2012; Thiene and Basso, 2015). Bausch-Jurken and Kotchen (2015) asserted that American Heart Association (AHA) had estimated the total cost, both direct and indirect, of cardiovascular disease (CVD) and stroke in the United States to be $312.6 billion. AHA also projected the cost of cardiovascular care to increase to an estimated $818.1 billion by 2030. AHA has attributed 40.6% of CVD to hypertension, 13.7% to smoking, 13.2% to poor diet, 11.9% to inactivity, and 8.8% to abnormal blood glucose. Walsh et al. 2014 reported physicians and healthcare networks have been slow to adopt electronic medical records and to integrate medical data with the ubiquitous mobile device. The need for cardiac diagnostics, like electrocardiography (ECG) holters or cardiac event recorder resulted in creation of such devices about 50-years ago (Wcislik et al. 2015). In cardiovascular prevention, there is classically a small number of cardiovascular risk factors to treat, such as hypertension, diabetes, hyperlipidemia and smoking excess, which are widely detected and treated. Recently, it has been widely recognized that new mechanical factors should be detected and treated and involves specifically pulsatile arterial hemodynamic (PAH) parameters such as: arterial stiffness, pulse pressure, and, to a lesser extent, augmentation index and pulse pressure amplification.

Mobile and wearable systems for continuous health monitoring are a key technology in helping the transition of health care to a more proactive and affordable healthcare. Wearable health monitoring systems allow an individual to closely monitor changes in her or his vital signs and provide feedback to help maintain an optimal health status. If integrated into a telemedical system, these systems can even alert medical personnel when life-threatening changes occur. Patients can benefit from continuous long-term monitoring as a part of a diagnostic procedure, can achieve optimal maintenance of a chronic condition, or can be supervised

during recovery from an acute event or surgical procedure (Milenković et al. 2006).

2. Literature Survey

Methods of medical diagnosis are continuously being improved and extended. Ulucam, 2012, identified the most well-known CVD risk factors in the elderly as high blood pressure (BP), wide pulse pressure, age (male > 55, women > 65), smoking, dyslipidemia (total cholesterol >190 mg/dL, or LDL cholesterol >115 mg/dL, or HDL cholesterol in men <40 mg/dL, female <46 mg/dL, triglyceride >150 mg/dL), fasting glucose 102-125 mg/dL, abnormal glucose tolerance test, diabetes mellitus, abdominal obesity (abdominal circumference: M > 102 cm, F > 88 cm), and a family history of premature CVD disease. Unhealthy lifestyle behaviours, including smoking, physical inactivity, hazardous alcohol consumption and low intake of fruit and vegetables have been shown to contribute to the development of coronary heart disease (CHD), which remains a leading cause of death worldwide (Dale et al. 2014). Kilty and Prentice, 2012 and Kong and Choi, 2012 reported CVDs have become one of the leading causes of morbidity and premature mortality in men and women in the industrialized world and many developing countries. The leading global risks for mortality in the world were high blood pressure (13% of global deaths), tobacco use (9%), high blood glucose (6%), physical inactivity (6%) and overweight or obesity. It was also predicted that by the year 2020, CVDs would be the leading cause of death in the entire world. Heart attacks and CHD are primarily caused by atherosclerosis, where a narrowing and hardening of the arteries result from an accumulation of fat and cholesterol deposits called plaque. Gaziano et al. 2015 documented cardiovascular disease (CVD) to have emerged as the single most important cause of death worldwide. In 2010, CVD caused an estimated 16 million deaths and led to 293 million disability-adjusted life-years (DALYs) lost — accounting for approximately 30% of all deaths and 11% of all DALYs lost that year. Like many high-income countries (HICs) during the past century, now low- and middle-income countries (LMICs) are seeing an alarming and accelerating increase in CVD rates.

Haslam and James, 2005 found CVD, with an emphasis on congestive heart failure, was being studied using proteomics and continues to be increasingly relevant to an aging population. Recently, NCD has become an important cause of mortality & morbidity in developing countries. Diabetes Mellitus (DM) and hypertension are major predisposing factors to CVD. Upsurge of DM & hypertension is propelled by growing prevalence of overweight and obesity worldwide, especially among children & adolescents. Heart failure, (HF), a condition where the heart is no longer able to maintain adequate blood circulation, results from myocardial dysfunction that impairs the heart's ability to circulate blood at a rate sufficient to maintain the metabolic needs of peripheral tissues and various organs. Heart failure is a relatively common clinical disorder, estimated to affect more than 2 million patients in the United States. About 400,000 new

patients develop congestive heart failure (CHF) each year. Morbidity and mortality rates are high; annually, approximately 900,000 patients require hospitalization for CHF, and up to 200,000 patients die from this condition. The average annual mortality rate is 40–50% in patients with severe (New York Heart Association (NYHA) class IV) heart failure (Deedwania, 2007). Some causes of heart failure include coronary artery disease, valvular disease, and myocardial infarction. Heart failure is a common disease in the Western world with a high prevalence and steadily rising incidence. Two major reasons contribute to the increasing incidence in this part of the world: firstly, better treatment of cardiovascular disease, in particular of acute ischemic events, such as myocardial infarcts which keep more people alive, however often at the cost of damaged, malfunctioning heart muscle, the first step on the road to heart failure; and secondly, an ageing population – heart failure is typically a disease of the elderly. The average age of the heart failure patient in the community is 74-75 years. Average prevalence of heart failure is 2-2.5% overall, increasing to >10% in octogenarians; up to 14 million inhabitants of Europe have heart failure; and average incidence of heart failure is 15/1000 inhabitants in people ≥55 years, but increases significantly in the elderly. It is now recognized that approximately 40% of all heart failure patients may have a preserved pump function of the left chamber (left ventricle- LV) of the heart. Patients with HF have a worse quality of life than those with almost any other chronic disease including bronchitis/emphysema, kidney failure and arthritis. Chronic diseases are common and costly, yet they are also among the most preventable health problems (WHFS, 2010).

2.1 Therapy and Treatment

reatment of cardiovascular disorders is one of the most highly evidence-based area of medicine and pharmacy practice. A careful patient history and physical examination are extremely important in diagnosing cardiovascular disease and should be done prior to any test. Heart sounds and heart murmurs are important in identifying heart valve abnormalities and other structural cardiac defects. Elevated jugular venous pressure is an important sign of heart failure and may be used to assess severity and response to therapy (Talbert, 2005). Accounting for more than 40% of deaths each year, cardiovascular disease remains the leading cause of mortality in the United States. Contributing to this mortality are two key conditions: myocardial infarction and congestive heart failure. Myocardial infarction triggers the formation of scar tissue, which is one of the causes of congestive heart failure. Current treatments for congestive heart failure provide a limited palliative outcome; therefore, myocardial infarction and congestive heart failure could benefit substantially from cell therapies. Such therapies could benefit not only patients but also the healthcare system in terms of burden of resources and financing (Sage, 2008). Improvements in health care and treatment of diseases have led to an increase in life expectancy in developed countries. However, this achievement has also inadvertently increased the prevalence of chronic illnesses such as

cardiovascular disease, adding to the growing burden of health care cost globally. Ironically, the recent improvements in treating ischemic disease have increased the number of patients living with congestive heart failure, the fastest-growing segment of cardiovascular disease. Unfortunately, this prevalence trend is expected to escalate in the foreseeable future. Cardiovascular disease remains one of the main problems in contemporary health care worldwide, accounting for approximately one third of the world's total death (Poole-Wilson, 2005). Although >80% of global burden of CVD occurs in developing countries, however, knowledge on the risk factors is largely derived from developed countries (Parvez, 2007). Kilty and Prentice (2012)'s model of CVD treatment as presented in Figure 1 reports that there is strong evidence that cardiovascular risk factors begin and can be identified in childhood and adolescence that influence the development of CVD in adulthood. They called for interdisciplinary and interprofessional teams of researchers, clinicians, educators, parents and care providers to work together on this health issue and inform each other of their outcomes.

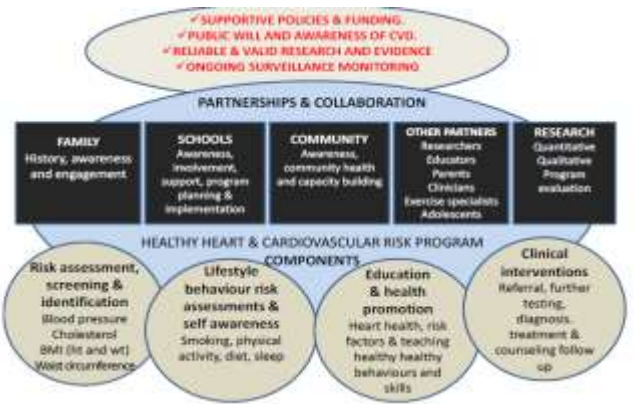

Figure 1 Comprehensive Treatment CVD Model (Kilty and Prentice, 2012)

New technologies are pertinent to generate high-dimensional data that provide unprecedented opportunities for unbiased identification of biomarkers that can be used to optimize pre-operative planning, with the goal of avoiding costly post-operative complications and prolonged hospitalization (Aggeli et al; 2014). Mobile technologies have been confirmed to offer the ability to connect patients with their doctors, care-givers and loved ones and enable timely health monitoring which suggests improved patient engagement and better health outcomes. Mobile technology provides aid in providing access to information, helping to lower costs, facilitating remote care and increasing efficiencies by connecting patients to their providers virtually anywhere. Mobile health applications and services are becoming an essential tool in extending health care resources around the world (West,

2013). Smart phone apps and wearable sensors are promising for improving cardiovascular health behaviors, preliminary data suggest. Self-monitoring is a key facet of changing behavior to prevent and manage heart health. Smartphone apps and wearable sensors have the potential to encourage positive change (AHA, 2015). Boursalie *et al* 2015 presented M4CVD: a Mobile Machine Learning Model for Monitoring Cardiovascular Disease, a system designed specifically for mobile devices that facilitates monitoring of cardiovascular disease (CVD). M4CVD using wearable sensors collects observable trends of vital signs contextualized with data from clinical databases. Instead of transferring the raw data directly to the health care professionals, M4CVD performs analysis on the local device by feeding the hybrid of collected data to a support vector machine (SVM) to monitor features extracted from clinical databases and wearable sensors to classify a patient as "continued risk" or "no longer at risk" for CVD. These statistics suggest that health care needs a major shift toward more scalable and more affordable solutions including measuring the rate of heartbeat using mobile computing technologies to monitor and ensure proper communication between the patient and cardiologist addressed in this paper.

2.2 Control of Cardiovascular System: Hemodynamic volumetric parameters

Hemodynamics has been defined as the study of the relationship among physical factors affecting blood flow through the vessels. Blood flow is a function of pressure difference and resistance. Blood flow (F) through a blood vessel is determined by two main factors: (1) pressure difference (ΔP) between the two ends of the vessel and (2) the resistance (R) to blood flow through the vessel (Figure 2).

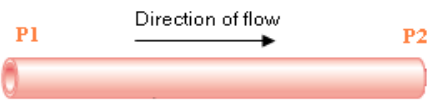

Figure 2 Blood flow through a blood vessel (Nasimi, 2012)

The equation relating these parameters is:

$$F = \Delta P/R \qquad (1)$$

This equation is called Darcy's law or Ohm's law.

Flow (F) is defined as the volume of blood passing each point of the vessel in one unit time. Usually, blood flow is expressed in milliliters per minute or liters per minute, but it is also expressed in milliliters per second. Pressure which is the force that pushes the blood through the vessel is defined as the force exerted on a unit surface of the wall of the tube perpendicular to flow.

Pressure is expressed as millimeters of mercury (mmHg). Since the pressure is changing over the course of the blood

vessel, there is no single pressure to use; therefore the pressure parameter used is pressure difference (ΔP), also called pressure gradient, which is the difference between the pressure at the beginning of the vessel (P1) and the pressure at the end of the vessel (P2), i.e. ΔP = P1 - P2. As seen in the Darcy's law, ΔP is the cause of the flow; with no pressure difference there would be no flow. The pressure energy is produced by the ventricle and it drops throughout the vessel due to resistance. In other words, resistance is the cause of the pressure drop over the course of a vessel. Resistance is how difficult it is for blood to flow from point 1 to point 2. Resistance impedes flow and it is a measure of interactions between flowing particles (including molecules and ions) themselves and interactions between flowing particles and the wall of the vessel.

As seen Darcy's law, resistance is the impeding cause of the flow; the bigger the resistance the lesser the flow. If the resistance is Δ (complete closure of the vessel) there will be no flow.

The resistance equation is:

$R = 8\eta L / \pi r4$ (2)

where η = fluid viscosity

L = vessel length

r = inside radius of the vessel.

Viscosity represents the interactions between flowing particles themselves and radius represents the interactions between flowing particles and the wall of the vessel. The units of viscosity are $Pa{\cdot}s = Ns/m^2$, or Poise ($dynes{\cdot}s/cm^2$), with 1 Pa·s = 10 Poise (Nasimi, 2012).

Rudenko et al. 2012 has asserted the foundation of hemodynamics as the phase mode of the heart performance such that in one beat, the heart changes its shape ten times that corresponds to the heart cycle phases. The most efficient way is to evaluate the status of hemodynamics not only by values of integral parameters, i.e., stroke and minute volumes, but also phase-related volumes of blood entering or leaving the heart in the respective phase in a cardiac cycle. The final formulae for calculating the volumes of blood in the phase of rapid and slow ejection, symbolized as PV3 and PV4, respectively, are as follows:

$PV3=S\bullet(QR+RS)^2 \bullet f_1(\alpha) \bullet (f_2(\alpha)+f_3(\alpha,\beta,\gamma,\delta)$ (ml); (1)

$PV4=S\bullet (QR+RS)^2\bullet f_1(\alpha)\bullet f_4(\alpha,\beta,\gamma,\delta)$ (ml), (2)

where S - cross-section of ascending aorta;

QR – phase duration according to ECG curve;

RS – phase duration according to ECG curve;

$f_1(\alpha)= 22072.5((5\alpha - 2)^3 - 27) / ((5\alpha - 2)^5 – 243);$

$f_2(\alpha)= (\alpha^5 – 1)/2;$

$$f_3(\alpha,\beta,\gamma,\delta)= \frac{1}{8}\left(\frac{10}{3}(4\alpha^2 -\delta^2)(\beta^3 -\alpha 3)\right.$$

$$\left.+5\chi\delta(\beta^4 -\alpha^4) - 2\chi^2(\beta^5 -\alpha^5)\right);$$

$$f_4(\alpha,\beta,\gamma,\delta)= \frac{1}{8}\left(\frac{5}{3}(\delta^2 - 8\alpha^2)(\beta^3 -\alpha^3) + \right.$$

$$\left.7.5\chi\delta(\beta^4 -\alpha^4) + 3\chi 2(\beta^5 -\alpha^5)\right);$$

$$\alpha = \frac{(1+ Em)^{0.2}}{QR+ RS};$$

$$\beta = \frac{(1+ Em+ Er)^{0.2}}{QR+ RS};$$

$\chi = 2(\alpha - 1) / (\beta – \alpha);$

$\delta =\alpha(2 +\chi).$

Stroke volume, SV, is calculated by an equation as given below:

$SV = PV3 + PV4 = S \bullet (QR+RS)^2 \bullet f_1(\alpha) \bullet (f_2(\alpha) + f_3(\alpha,\beta,\gamma,\delta) + f_4(\alpha,\beta,\gamma,\delta))$ (ml) (3)

The minute stroke is computed as follows:

$MV = SV \bullet HR$ (l/min) (4)

In similar way calculated are other phase-related volumes of blood as listed below:

PV1 – volume of blood entering the ventricle in premature diastole;

PV2 – volume of blood entering the ventricle in atrial systole;

PV5 – volume of blood pumped by ascending aorta as peristaltic pump.

So, the main parameters in hemodynamics are 7 volumes of blood entering or leaving the heart in different heart cycle phases. They are as follows: stroke volume SV, minute volume MV, two diastolic phase-related volumes PV1 and PV2, two systolic phase-related volumes PV3 and PV4, and PV5 as volume of blood pumped by the aorta. These hemodynamic parameters should be used mainly in order to evaluate eventual deviations from their normal values, if any. The limits of normal values of hemodynamic parameters are not conditional, and they have their respective calculated values. With respect to the normal values (the required parameters) in hemodynamics, they have been taken on the basis of the known data on ECG waves, intervals and segments for adults from the literature sources as given below:

1. The upper and lower limit of the QRS complex values:

$QRS_{max} = 0.1$ s;

$QRS_{min} = 0.08$ s.

2. The upper and lower limit of the RS complex values:

$RS_{max} = 0.05$ s;

$RS_{min} = 0.035$ s.

3. The normal value of interval QT in every specific cardiac cycle is determined from the Bazett formula as follows:

$QT = 0.37 \sqrt{RR}$ 0.5 s (for men);

$QT = 0.4 \sqrt{RR}$ 0.5 s (for women).

4. Normal value PQ is calculated from a formula as indicated below:

$PQ = 1 / (10^{-6} \; 638,44 \; HR^2 + 9,0787)$ s.

This equation has been produced according to the method of approximation of normal values PQ, as known from the sources, considering their dependence on heart rate (HR) (Rudenko et al. 2012).

Hemodynamic instability is most commonly associated with abnormal or unstable blood pressure (BP), especially hypotension, or more broadly associated with inadequate global or regional perfusion. Inadequate perfusion may compromise important organs, such as heart and brain, due to limits on coronary and cerebral auto regulation and cause life-threatening illnesses, or even death. Therefore, it is crucial to identify patients who are likely to become hemodynamically unstable to enable early detection and treatment of these life-threatening conditions (Cao et al. 2008). Modern intensive care units (ICU) employ continuous hemodynamic monitoring (e.g., heart rate (HR) and invasive arterial BP measurements) to track the state of health of the patients. However, clinicians in a busy ICU would be too overwhelmed with the effort required to assimilate and interpret the tremendous volumes of data in order to arrive at working hypotheses. Consequently, it is

important to seek to have automated algorithms that can accurately process and classify the large amount of data gathered and to identify patients who are on the verge of becoming unstable (Cao *et al*. 2008). Modern ICUs are equipped with a large array of alarmed monitors and devices which are used to try to detect clinical changes at the earliest possible moment so as to prevent any further deterioration in a patient's condition. The effectiveness of these systems depends on the sensitivity and specificity of the alarms, as well as on the response of the ICU staff to the alarms. However, when large numbers of alarms are either technically false, or true, but clinically irrelevant, response efficiency can be decreased, reducing the quality of patient care and increased patient (and family) anxiety (Nataraj *et al*. 2012).

3. Statement of the Problem

Heart Failure (HF) is a leading cause of hospitalization for people 65 years of age and older, and rates of hospital readmission within 6 months range from 25% to 50%. HF is managed by patients suffering from it visiting the doctor regularly for check up and treatment. Patients are not fully involved in some vital tasks, which if they could do for themselves would ease the doctors of some work.

4. Methods

We administered 485 questionnaires and interviewed 12 cardiologists, 45 physicians, and 23 pharmacists. Pertinent questions during the data collection phases centred on cardiovascular parameters including Blood Viscosity, Cholesterol Crystal, Blood Fat, Vascular Resistance, Vascular Elasticity, Myocardial Blood Demand, Myocardial Blood Perfusion Volume, Myocardial Oxygen Consumption, Stroke Volume, Left Ventricular Ejection Impedance, Left Ventricular Ejection Impedance, Left Ventricular Effective Pump Power, Coronary Artery Elasticity, Coronary Perfusion Pressure, Cerebral Blood Vessel Elasticity, and Brain Tissue Blood Supply Status. Cognate registries including Cardiac Arrest Registry to Enhance Survival (CARES), the Cardiovascular Research Network (CVRN), the National Cardiovascular Data Registry (NCDR), the International Registry of Aortic Dissection (IRAD), and the Global Registry for Acute Cardiac Events (GRACE) were consulted to collect information on cardiovascular disease. Literature databases such as MEDLINE, APAIS, Google Scholar and the Clinicians Health Channel were searched. Search terms used included "cardiovascular*", "mortality*", "cardiac", "heart*", "blood*", "non-communicable*", "hyperten*", "myocardial*", and "risk factor".

As guided by international standards of the Institute of Medicine (IOM), detailed information on chronic conditions—including cardiovascular disease, diabetes, and respiratory health and disease—were collected by the administered questionnaire, and participants were assisted to undergo comprehensive dietary interviews and body measurements. The cardiologists, by standard practice, undertook physical examination that included several measures relevant to CVD and respiratory diseases, including blood pressure and spirometry, as well as

cardiovascular fitness, body mass index, and body composition. Relevant biomarkers include cholesterol and triglyceride measures, C-reactive protein, and fasting plasma glucose. In addition to interviews with cardiologists to gather cognate questions, this project employed Unified Modeling Language (UML)'s use case, sequence, collaboration diagrams to formalize the functional requirements / interaction between a patient and a cardiologist as shown in Figure 3.

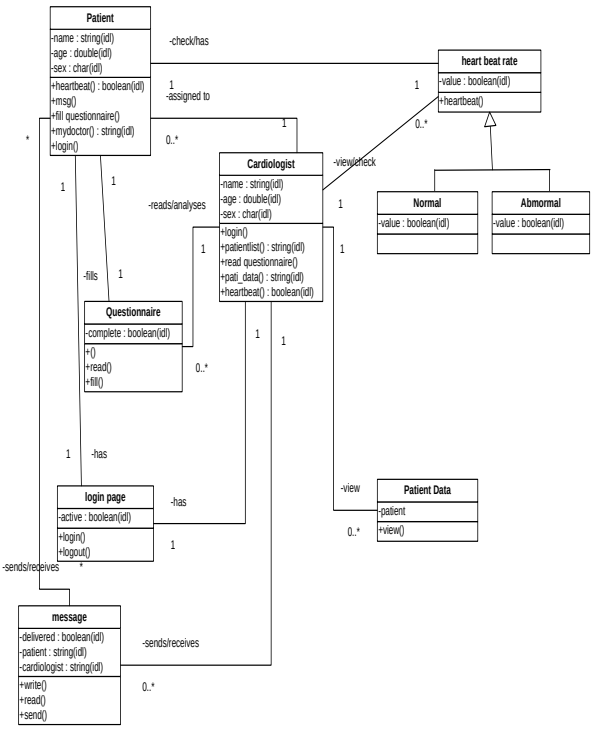

Figure 3 System Class Diagram of the CVDMS

The resulting framework was implemented on Edition Java 2 Platform (J2ME), MySQL Server 5.0, Java servlets, Apache Tomcat 6.0 server, and Ozeki sms server for emergency sms. Cardiovascular Diseases Management System (CVDMS) has modules designed for the patient's end to aid proper monitoring by the cardiologist and proper communication with the cardiologist.

5. Results

469 of 485 questionnaires (96.70%) were validly completed and returned, while 16 (3.30%) were not. Cardiovascular parameters normal range values (lower bound, median, and upper bound) confirmed from cardiologists are as presented in Figure 4.

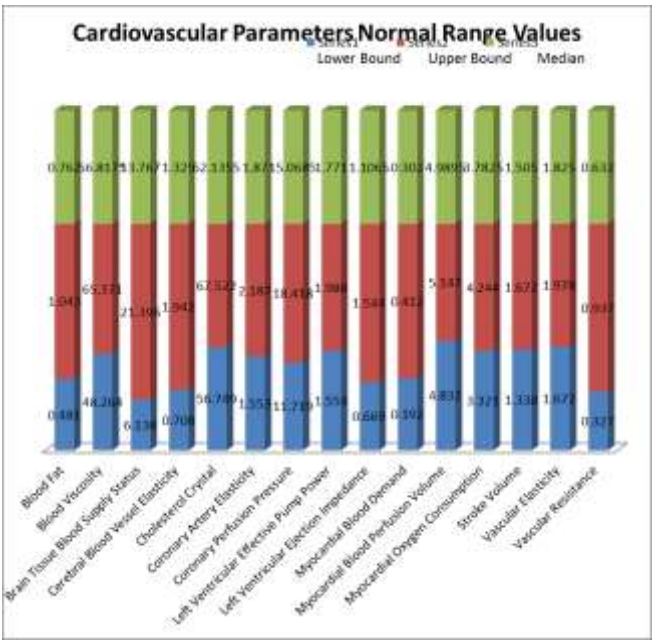

Figure 4 Cardiovascular Parameters Normal Range Values

Cardiologists confirmed, among other pertinent things, that Myocardial Oxygen Consumption (the milliliter value of oxygen consumption of heart per minute) is influenced by: (1) Heart rate: the heart rate is fast, and the HOV is great; (2) Myocardial contractility: the cardiac contractility is strong, and the HOV is great; and (3) Myocardial contraction time: the longer the contraction time is, the greater the HOV is. Thus, low oxygen consumption and high cardiac work are the best state. High blood pressure patients with high viscosity are prone to have cerebrovascular accidents, such as stroke and other phenomena; coronary heart disease patients with high viscosity are prone to have myocardial infarction and so on. Increase is in direct proportion to the length of blood vessels, and is in inverse proportion to the caliber of blood vessels. The increase of vascular resistance is seen in mildly elevated systolic and diastolic blood pressure, mild hypertension, insomnia with deficiency of heart and spleen, phlegm-heat internal confusion type insomnia, etc. Decline is seen in mildly declined systolic and diastolic blood pressure, mild hypotension, Yin deficiency and Huo exuberance type insomnia, etc. In a case of a 59year-old male, 85kg and 175cm height, the measurements collected were as shown in Figures 5, 6, and 7 to determine the risk level (severe partial fat).

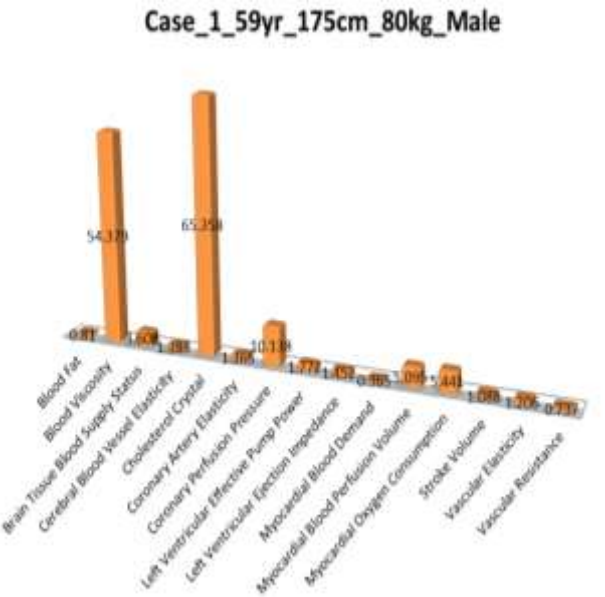

Figure 5 Cardiovascular Parameters actual measurements for a 59-yr-old, 175cm, 80kg male

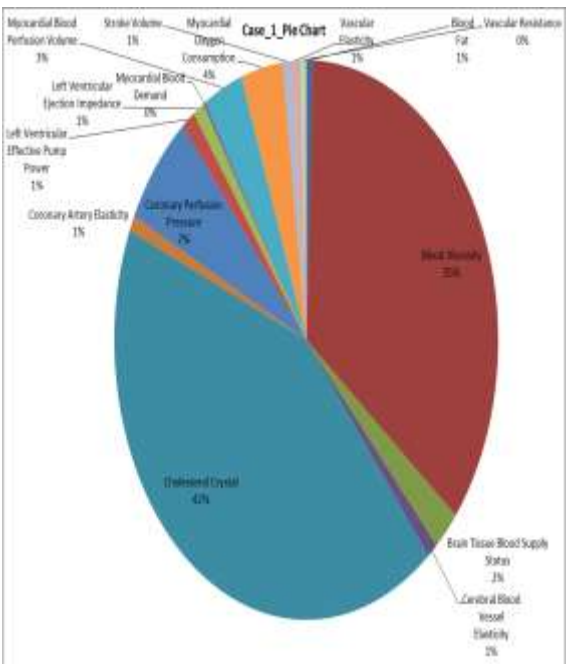

Figure 6 Pie Chart of Cardiovascular Parameters actual measurements for a 59-yr-old, 175cm, 80kg male

(Cardiovascular and Cerebrovascular)
Analysis Report Card

Name: Omeghadegun Sex: Male Age: 50
Figure: Severe partial fat(175cm,80kg) Testing Time: 2012-03-03 15:02

Actual Testing Results

Testing Item	Normal Range	Actual Measurement Value	Testing Result
Blood Viscosity	48.264 - 65.371	54.379	Normal (-)
Cholesterol Crystal	56.749 - 67.522	63.338	Normal (-)
Blood Fat	0.481 - 1.043	0.81	Normal (-)
Vascular Resistance	0.327 - 0.937	0.735	Normal (-)
Vascular Elasticity	1.672 - 1.978	1.200	Moderately Abnormal (++)
Myocardial Blood Demand	0.192 - 0.412	0.365	Normal (-)
Myocardial Blood Perfusion Volume	4.832 - 5.147	5.089	Normal (-)
Myocardial Oxygen Consumption	3.321 - 4.244	5.441	Mildly Abnormal (+)
Stroke Volume	1.338 - 1.672	1.988	Mildly Abnormal (+)
Left Ventricular Ejection Impedance	0.669 - 1.544	1.452	Normal (-)
Left Ventricular Effective Pump Power	1.554 - 1.988	1.772	Normal (-)
Coronary Artery Elasticity	1.553 - 2.187	1.765	Mildly Abnormal (+)
Coronary Perfusion Pressure	11.719 - 18.418	10.138	Moderately Abnormal (++)
Cerebral Blood Vessel Elasticity	0.708 - 1.942	1.383	Normal (-)
Brain Tissue Blood Supply Status	0.138 - 21.396	3.608	Mildly Abnormal (+)

Reference Standard: **- Normal + Mildly Abnormal ++ Moderately Abnormal +++ Severely Abnormal**

Blood Viscosity: 48.264-65.371(-) 65.371-69.645(+) 69.645-73.673(++) >73.673(+++)

Cholesterol Crystal: 56.749-67.522(-) 67.522-69.447(+) 69.447-74.927(++) >74.927(+++)

Blood Fat: 0.481-1.043(-) 1.043-1.669(+) 1.669-1.802(++) >1.802(+++)

Vascular Resistance: 0.327-0.937(-) 0.937-1.543(+) 1.543-1.857(++) >1.857(+++)

Vascular Elasticity: 1.672-1.978(-) 1.672-1.511(+) 1.511-1.047(++) <1.047(+++)

Myocardial Blood Demand: 0.192-0.412(-) 0.412-0.573(+) 0.573-0.709(++) >0.709(+++)

Myocardial Blood Perfusion Volume: 4.832-5.147(-) 4.177-4.832(+) 4.029-4.177(++) <4.029(+++)

Myocardial Oxygen Consumption: 3.321-4.244(-) 4.244-5.847(+) 5.847-6.472(++) >6.472(+++)

Stroke Volume: 1.338-1.672(-) 0.647-1.338(+) 0.339-0.647(++) <0.339(+++)

Left Ventricular Ejection Impedance: 0.669-1.544(-) 1.544-2.057(+) 2.057-2.417(++) >2.417(+++)

Left Ventricular Effective Pump Power: 1.554-1.988(-) 1.076-1.554(+) 0.507-1.076(++) <0.507(+++)

Coronary Artery Elasticity: 1.553-2.187(-) 1.183-1.553(+) 0.983-1.182(++) <0.983(+++)

Coronary Perfusion Pressure: <8.481(+++) 8.481-11.719(++) 18.418-21.274(+++) >21.274(+++)

Cerebral Blood Vessel Elasticity: 0.708-1.942(-) 0.453-0.708(+) 0.309-0.453(++) <0.947(+++)

Brain Tissue Blood Supply Status: 0.138-21.396(-) 3.219-0.138(+) 21.274-27.259(+++) >27.214(+++)

Figure 7 cardiovascular and Cerebrovascular Analysis Report Card

We also gathered from the cardiologists that stroke volume (the blood volume output by the heart in beat each time) are equally influenced by: (1) The effective circulating blood volume (BV): when the blood volume is insufficient, the returned blood volume is little, and the SV is reduced; (2) The weakening of myocardial contractility: the contractility is low, and the pressure is low, so the ejected blood volume is less; (3) The extent of ventricular filling: In range of myocardial elasticity, the greater the degree of filling is, the stronger the retraction is, and the SV is increased. The normal heart chamber capacity is 173ml, but not all of the blood is ejected. The blood volume in the left ventricle is about 60% -70% of the total capacity, being about 125ml or so; (4) The size of peripheral vascular resistance (PR). The PR is large, and then the SV is reduced; the PR is small, and then the SV is increased; and (5) Ventricle wall movement. When the ventricle is contracted, the cardiac muscle is in coordinated movement. If the myocardial contraction is not coordinated, the SV is reduced. For instance, some patients with myocardial infarction have part of infarction, so the myocardial contractility is inconsistent and the SV is reduced. However, under normal circumstances, the ventricle wall movement can not be abnormal.

Figure 8 presents the login module for proper authentication of the user of this application, precisely the patient. The patient is given a list of options specifying the various functions that can be performed by the application on the patient's end.

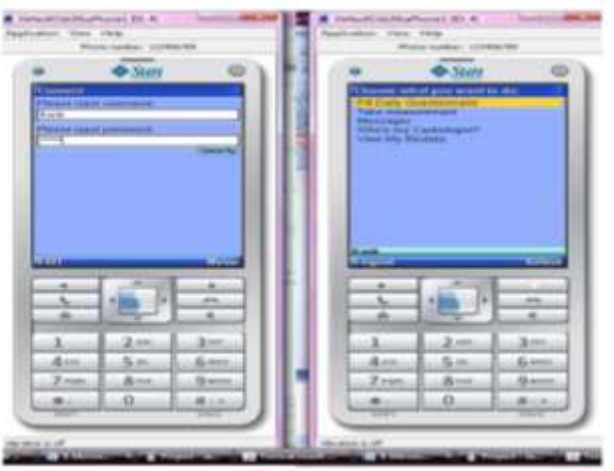

Figure 8 Login Menu module

As part of the cardiologist's monitoring exercise, he needs to have a daily report on the patient's health. As such, this module enables the patient fill a questionnaire daily as shown in Figure 9 in order to keep the cardiologist abreast of the patient's health status. The questions to be filled are basic general questions that help doctors in determining the general state of the patient's heart.

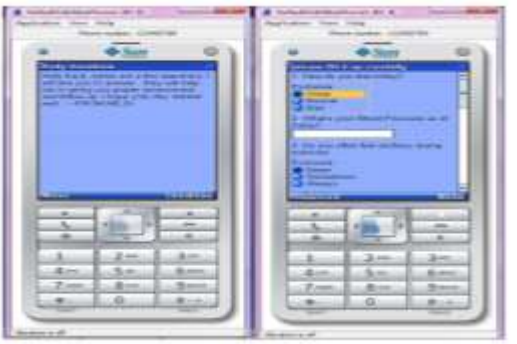

Figure 9 Daily Questionnaire & its Filling

After filling the questionnaire, a confirmation screen is displayed as shown in Figure10.

Figure 10 Screenshot showing the confirmation screen

Figure 11 provides help for the patient to send and receive vital messages to and from the cardiologist. This is also needed for proper monitoring of a patient and as such management of the cardiovascular disease.

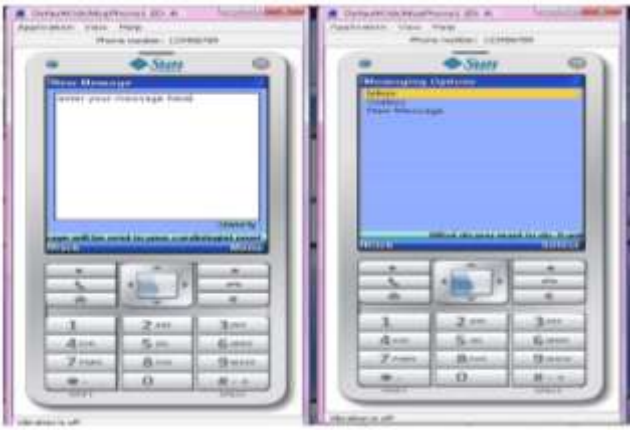

Figure 11 New Message screen / Messaging Option

A patient can view his message inbox (messages sent by the doctor to the patient), his message outbox (messages sent from the patient to the cardiologist). The CVDMS also provides an avenue for sending messages pertaining to health issues to the cardiologist.

The *Take Measurement Module* incorporates the Bluetooth technology to receive the rate of the patient's heartbeat from the CVDMS heart monitoring device as shown in Figure 12.

The device which acts as a slave finds the mobile phone and the service it offers, then sends the data to the mobile phone which acts as the master. The Java Bluetooth API plays an important role here as it enables better and easy communication between both Bluetooth devices.

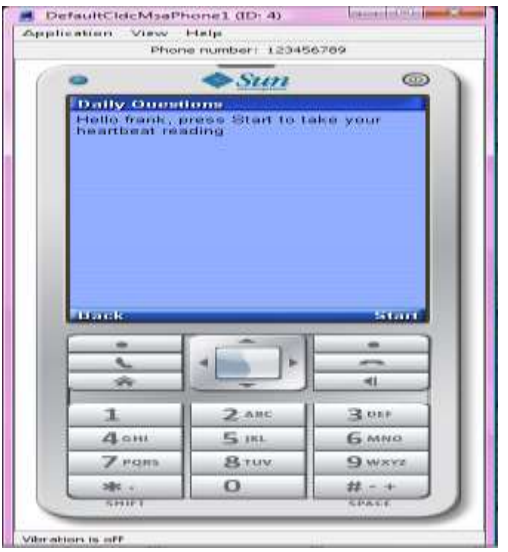

Figure 12 Screenshot of the introduction to commence heartbeat reading

After taking the measurement, a java servlet is called to determine if the heart rate is within the normal range. If not, a message is stored in the ozekimessageout table and tagged as "send". This means that the message is pending. The Ozeki SMS Gateway is configured to check the ozekimessageout table every 5 seconds to check for pending messages. In case of pending messages, the server sends an emergency sms to the phone of the next of kin, the cardiologist and the hospital. This will ensure proper monitoring and management of the cardiovascular disease. The patient can view the biodata of both himself and his cardiologist's as shown in Figure 13.

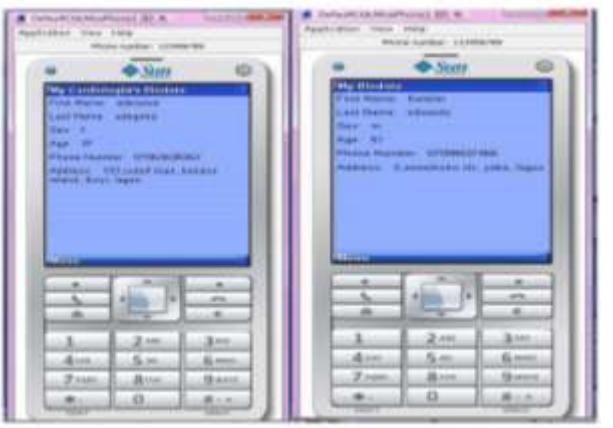

Figure 13 Screen shots of Cardiologist's & Patient's Biodata

5.1 The CVDMS Monitoring Device

In order to take proper reading and measurement of the heartbeat, a microcontroller was used for processing and an output device called the Liquid Crystal Display(LCD) was used to display the heartbeat rate. The signals sent to the green LED, an indicator for the heartbeat, was sent to an STC 8051 microcontroller and the pulses were counted within a space of one minute so as to know the rate of heartbeat per minute. After determining the rate, the value is then displayed on the LCD. This was first simulated using the ISIS 7 Professional and the result's screenshot is shown in Figure 14.

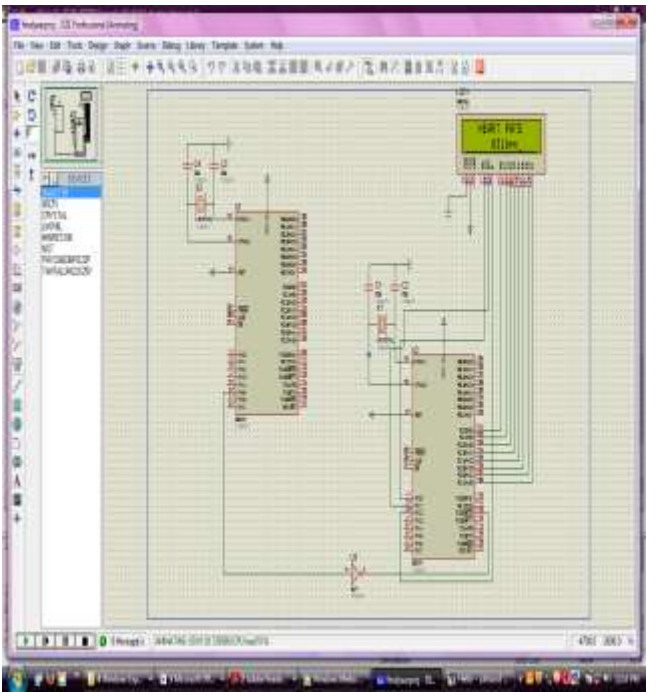

Figure 14 Simulation of microcontroller interfaced with LCD (CVDMS heart monitor)

When the mobile application gets the heartbeat reading using its Bluetooth technology, if the heart rate is abnormal, a trigger is set to enable the Ozeki SMS Gateway send an alert to the patient's next of kin and cardiologist. The obtained values could be substituted in Rudenko et al. 2012's equations to assist the cardiologist in his decision-making.

6. Conclusion and future work

Chronic diseases have been adjudged as a costly part of current healthcare delivery system as nearly three-quarters of medical expenditures have been recorded to have taken place on a small number of chronic illnesses, including cardiovascular disease, cancer, diabetes, and asthma. Heart failure is the cause of a high rate of readmission and it ultimately leads to death if not properly managed and supervised, thereby making cardiovascular disease remain one of the main problems in contemporary health care worldwide, accounting for approximately one third of the world's total death. The growing incidence of diabetes mellitus and the continuing epidemic of cardiovascular disease associated with this ailment have induced numerous investigators to seek evidence of pre-clinical disease besides trying to diagnose advanced stages of disease. Using a novel smartphone adapter, patients are now able to capture and transmit single-lead ECG data to their healthcare providers. Consequently, remote patient monitoring has increasingly become an attractive solution for the management of CVD. This paper, through mobile computing technologies, has succeeded in achieving acquisition of biological

signals (heartbeat) and make them available wirelessly over Bluetooth. This allows an individual patient's direct involvement to closely monitor changes in her or his vital signs and provide feedback to help maintain an optimal health status. Patients and care providers can both benefit from remote monitoring as it helps patients be more engaged in their health through self-reported outcomes and provides support for cost-effective care. It also alerts medical personnel when life-threatening changes occur thereby ensuring proper communication between the patient and cardiologist via messaging towards reducing incidence of re-admission. An accurate assessment of BP levels and early identification and treatment of hypertension is thus essential for reducing the cardiovascular risk associated with this condition. The use of mobile systems that monitor patient symptoms and provide real-time advice on treatment and medication because they have the potential to control costs, reduce errors, and improve patients' experiences should be encouraged. The Cardiovascular Disease Management System (CVDMS), will be evaluated by its accuracy in classifying live monitored data. We will continue to explore methods to test the system's sensitivity to changing patient conditions towards the system's improvement following ubiquity of technology.

References

Aggeli, Constantina; Felekos, Ioannis; Poulidakis, Emanuel; Koutagiar, Iosif; Venieri , Erifili; and Stefanadis, Christodoulos (2014). Coronary Artery Bypass Grafting and Biomarkers. In Tousoulis, Dimitris and Stefanadis, Christodoulos (eds), *Biomarkers in Cardiovascular Diseases*, CRC Press, pp. 164-209.

AHA - American Heart Association (2015). Mobile technology may help people improve health behaviors, https://www.science-daily.com/ releases/2015/08 /150813171201.htm. Accessed 2 September, 2016.

Al-Jaafreh, Moha'med O. and Adel A. Al-Jumaily (2008). Blood Pressure Estimation with Considering of Stroke Volume Effect. In Nilmini Wickramasinghe and Eliezer Geisler (Eds): *Encyclopedia of Healthcare Information Systems, Medical Information Science Reference*, New York, 2008. ISBN 978-1-59904-889-5

Bausch-Jurken, Mary T. and Kotchen, Theodore A (2015). Risk Factors for Hypertension and Cardiovascular Disease. In Peplow ,Philip, Adams, James; and Young, Tim (eds),*Cardiovascular and Metabolic Disease: Scientific Discoveries and New Therapies*, pp. 137 - 159

Boursaliea, Omar; Samavia, Reza; and Doyle, Thomas E (2015). M4CVD: Mobile Machine Learning Model for Monitoring Cardiovascular Disease. The 5th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH 2015), *Procedia*

Computer Science 63 (2015), 384 – 391

Cao, H., Eshelman, L; Chbat, N; Nielsen, L; Gross, B; and Saeed, M (2008). Predicting ICU hemodynamic instability using continuous multiparameter trends, *Conf Proc IEEE Eng Med Biol Soc*, pp. 3803-3806, Vancouver, Canada, August 21-23, 2008

Dale, Leila Pfaeffli; Whittaker, Robyn; Jiang, Yannan; Stewart, Ralph; Rolleston, Anna ; and Maddison, Ralph (2014). Improving coronary heart disease self-management using mobile technologies (Text4Heart): a randomised controlled trial protocol, Trials 2014, 15:71. doi:10.1186/1745-6215-15-71

Deedwania, Prakash C. and Carbajal, Enrique V (2007). Congestive Heart Failure. In Crawford, Michael H (ed.), *Current Diagnosis and Treatment in Cardiology*, second edition, Chapter 18, The McGraw-Hill Companies.

Gaziano, Thomas A; Prabhakaran, Dorairaj; and Gaziano, J. Michael (2015). Global Burden of Cardiovascular Disease. In Mann, Douglas L; Zipes, Douglas P; Libby, Peter; Bonow, Robert O; Goldberg; Max and Lilly; and Braunwald, Eugene (eds), *Braunwald's Heart Disease: A Textbook Of Cardiovascular Medicine*, 10th Edition, Elsevier, pp. 23-42.

Haslam, D.W. and James, W.P. (2005). Obesity. *Lancet* 2005; 366:1197-1209

Kilty, Heather Lee and Prentice, Dawn (2012). Early Identification of Cardiovascular Risk Factors in Adolescents and Follow-Up Intervention Strategies. In Gasparyan, Armen Yuri (ed.), *Cardiovascular Risk Factors*, InTech, Croatia, pp. 25 – 68.

Kong, Alice P.S. and Choi, Kai Chow (2012). Novel and Traditional Cardiovascular Risk Factors in Adolescents. In Gasparyan, Armen Yuri (ed.), *Cardiovascular Risk Factors*, InTech, Croatia, pp. 69 – 88

Milenković, Aleksandar; Otto, Chris; and Jovanov, Emil (2006). "Wireless Sensor Networks for Personal Health Monitoring: Issues and an Implementation", 2006.

Nasimi, Ali (2012). Hemodynamics. In Gaze, David (ed.), *The Cardiovascular System - Physiology, Diagnostics and Clinical Implications*, pp. 108-121. InTech.

Nataraj, C; Jalali , A; and Ghorbanian, P (2012). Application of Computational Intelligence Techniques for Cardiovascular Diagnostics . In Gaze, David (ed.), *The Cardiovascular System - Physiology, Diagnostics and Clinical Implications*, pp. 224-253. InTech.

Parvez, H; Bisher, K; and El, Nahas M (2007). Obesity and Diabetes in the developing world – a growing challenge. *New England Journal of Medicine* 2007; 356: 213-215

Poole-Wilson, P (2005). The prevention of cardiovascular disease worldwide: Whose task and

WHO's task? *Clinical Medicine*, 5(4), 379-384.

Rudenko, Mikhail; Voronova, Olga; Zernov, Vladimir; Mamberger, Konstantin; Makedonsky, Dmitry; Rudenko, Sergey; Fedossov, Yuri; Duyzhikov, Alexander; Orlov, Anatoly; and Sobin, Sergey (2012). Control of Cardiovascular System. In Gaze, David (ed.), *The Cardiovascular System - Physiology, Diagnostics and Clinical Implications*, pp. 3-22; InTech.

Sage (2008). *The SAGE Handbook of Healthcare: Global Policies ● Business Opportunities ● Scientific Developments*, Decision Resources Inc, SAGE Publications Ltd, London

Talbert, Robert L (2005). Cardiovascular Testing. In Joseph T. DiPiro, Robert L. Talbert, Gary C. Yee, Gary R. Matzke, Barbara G. Wells, L. Michael Posey (eds), *Pharmacotherapy: Pathophysiologic Approach*, Sixth Edition MCGRAW-HILL Medical Publishing Division New York

Thiene, Gaetano and Basso, Cristina (2015). Foreward II. In Lucena, Joaquín S.; García-Pavía, Pablo; Suárez-Mier, M. Paz; and Alonso-Pulpón, Luis A (eds), *Clinico-Pathological Atlas of Cardiovascular Diseases*, Springer

Ulucam, Melek Z (2012). Cardiovascular Risk Factors in the Elderly. In Gasparyan, Armen Yuri (ed.), *Cardiovascular Risk Factors*, InTech, Croatia, pp. 89 – 110

Walsh III, Joseph A; Topol, Eric J; and Steinhubl, Steven R (2014). Novel Wireless Devices for Cardiac Monitoring, Circulation. 2014;130:573-581. DOI: 10.1161/CIRCULATIONAHA.114 .009024

Wcislik, M; Pozoga, M; and Smerdzynski, P (2015). Wireless Health Monitoring System, *IFAC-PapersOnLine* 48-4 (2015) 312–317

West, Darrell M. (2013). Improving Health Care through Mobile Medical Devices and Sensors, Center for Technology Innovation at Brookings, Washington, DC 20036

WHFS – World Heart Failure Society (2010). Heart Failure Worldwide - Facts and Figures, www.whfs.org

Appendix

List of abbreviations used in this manuscript:

BP	Blood Pressure
CAD	Coronary artery disease
CBC	Complete Blood Count
CHD	Coronary Heart Disease
CHF	CONGESTIVE HEART FAILURE
CVA	Cerebrovascular accident
CVD	Cardiovascular Disease

DM Diabetes Mellitus
HF Heart Failure
HTML HyperText Markup Language
MDG Millennium Development Goal
MI Myocardial infarction
MySQL Microsoft Structured Query Language
NCD Non-Communicable Disease
NYHA New York Heart Association
PHP Hypertext Processor Processing Language
PWV Pulse Wave Velocity
UML Unified Modelling Language
WHO World Health Organization

Acknowledgment

Dr. Eloho Edosio (a cardiologist, University of Lagos Teaching Hospital (LUTH), Lagos, Lagos State),

Dr. Godwin Adebose Olawale (Physician, Public & Reproductive Health, Ministry of Health, Akure, Ondo State),

Dr. Michael Adeboro Alabi (Physician, St. Michael Medical Centre, Akure, Ondo State),

Dr. Funsho Oladipo (Physician, RJolad Hospital Nig. Ltd, Bariga, Lagos), Dr. Toogun (Physician),

Engineer Reuben Olanipekun Aladetoyinbo (Director, Ministry of Agriculture, Akure, Ondo State - posthumously);

Professor Adetokunbo Babatunde Sofoluwe (Professor of Computer Sciences & Vice-Chancellor, University of Lagos, Lagos - posthumously); ***Professor Charles Onuwa Uwadia*** (Professor of Computer Sciences, University of Lagos, Lagos);

Professor Louis Osayenum Egwari (Professor of Biological and Medical Sciences Research, Covenant University, Ota, Ogun State);

Professor Victor W. Mbarika (Professor of Management Information Sciences & Healthcare Informatics Research, Southern University and A&M College, Baton Rouge, Louisiana, USA);

Chief Pius Oluwole Akinyelure (Idanre, Ondo State);

Dr. (Mrs) Mary Adeyanju (Registered Nurse, Diabetes / HIV Educator, and Director of Nursing Services Department, Ministry of Health, Ado-Ekiti, Ekiti State),

Mrs Chikaodili Amalachi Ukegbu (Pharmacist, The Federal Polytechnic Medical Centre, Ado-Ekiti, Ekiti State),

Miss Oluwayemisi 'Tosin Oluwasusi (Registered Nurse, Government State Hospital, Ado-Ekiti, Ekiti State),

Abiola Owoniyi, Mr. and Mrs. Abiodun, Global Health Workforce Alliance (GHWA), and Canadian Coalition for Global Health Research (CCGHR).

Covenant Journal of Informatics & Communication Technology Vol. 4 No. 2, Dec, 2016

An Open Access Journal, Available Online

Analysis of Modified Rule Extraction Algorithm and Internal Representation of Neural Network

Vinita Srivastava[1] & Chitra Dhawale[2],

[1]Department of I.I.C.C. , R.T.M. Nagpur University, Nagpur India.
[1]vinni.sara12@gmail.com
[2]Department of Computer Science, Amravati India.
[2] cadhawale@rediffmail.com

Abstract: Classification and Rule extraction is an important application of Artificial Neural Network. To extract fewer rules from multilayer feed forward neural network has been a research area. The internal representation of the network is augmented by a distance term to extract fewer rules from the feed forward neural network and experimented on five datasets. Understanding affect of different factors of the dataset and network on extraction of a number of rules from the network can reveal important pieces of information which may help researchers to enhance the rule extraction process. This work investigates the internal behavior of neural network in rule extraction process on five different dataset.

Keywords: Rule extraction, Feed Forward Neural Network, Hidden units, Activation value, Hidden neurons.

1. Introduction

Classification and pattern recognition is one of the important application of Artificail Neural Network [Kamruzzan et al. 2011][Sheng and Qi 2011]. Classification rules, extracted by efficient rule extraction algorithm helps in decision process but concepts learned by neural networks are difficult to understand because they are represented using large assemblages of real valued parameters [Srivastava et al.,2015]. A neural network is trained on a training input sample. A multilayered network has more than two layers. The layer between input layer and output layer is called as hidden layer and hence neurons in hidden layer are known as hidden neurons. Each layer is connected to each other and each connection has a weight associated with it. Weights and

biases are initialized with random values [Mia et al.,2015]. The training sample of input is given to the input layer, which gives an output of the layer by applying transfer function and weight to the input value. The output is then presented as input to hidden layer having its own transfer function, weights and biases. In this way the output of the hidden layer, called as the activation value of hidden neurons, is then presented to output layer which in turn gives the output of the network on the input sample. The calculated output is then compared with the target output. The difference between the two is calculated. The weights of the network are then changed by a small value, calculated by the training algorithm, to minimize the difference between the actual output and target output known as an error [Kamruzzaman & Hassan,2005].

The most minimized value of the error term is known as best performance value. After training and validation, network is trained to accept input samples and give the expected output value, i.e. value of selector attribute for that input sample.

A network is trained by a training algorithm which works on the principle of minimizing the error term. The error term is calculated at output layer by squaring the difference between the target output and the actual output. The goal of the training algorithm is to minimize this error term by changing representation of hidden units at each iteration, which makes it complex and needs more rules to explain. Researchers and scholars have worked on improvisation on training algorithms but not much on the internal representation of a network, the hidden layers and weights calculated at hidden neurons in comparison to training algorithms. Since in decompositional approach of rule extraction, rules are extracted from hidden units, therefore the number of rules mainly depends on hidden units and internal layer representation. During training hidden unit activation values can take their values anywhere in the space to achieve minimum squared error term calculated by the learning algorithm. Hence to get fewer rules, it is important to minimize the scattered activation values at hidden units in the network.

Rule extraction algorithm extracts rules from the trained network in terms of input and output [Kamruzzaman & Sarkar,2011]. It express the symbolic rules, for example, if for an input sample of a patient, (x1,x2,x3,x4,x5,x6,x7,x8,x9,x10) represents the values of 10 attributes of a patient then rules are expressed as:

If x1 (relational operator) v1 and x2 (relational operator) v2 and x3 (relational operator) v3 ……. and x10 (relational operator) v10 then

The patient will be a liver patient where vi is a numeric value for that attribute.

Since an input data sample can have many combinations of the values of attributes, the generalized number of rules should be minimum for taking decision on given set of values of attributes that whether the sample will be a patient or not, without affecting the accuracy of the decision.

The proposed rule extraction method, efforts to minimize the number of rules extracted from the network without affecting the classification accuracy.

2. The Distance Term

The three layered feed forward network is simulated and trained on the input samples.

The proposed method [Srivastava et al.,2015] follows the decompositional approach of rule extraction in which first rules are extracted between hidden units and output and then between input and hidden units. Combining the two gives rules in terms of input and output [Huynh & Reggia, 2011]. Since the number of rules mainly depends on hidden units and internal layer representation, the proposed method attempts to extract a fewer rule at hidden layer. For an input sample xp, the activation value of a hidden unit Hi is represented as apHi and actual output as op and target output as tp.

The activation values of hidden units will range from 0 to 1 after the logarithmic transfer function and target output will be either 0 or 1. For N hidden units, the activation values will be apH1 to apHN. The proposed method, calculates the Euclidean distances between hidden unit's activation value. For a given input, if the difference between two hidden unit's activation value is more, it means the two hidden units tends to give different output values for the same input. The activation value which is not clustered with other hidden unit's activation value will have a larger distance value from them in comparison to others (not clustered with others, but is in proximity), which represents that this hidden unit is tending to an intermediate output value. Such activation value is pushed towards the clustered hidden units to eliminate the rule with that intermediate hidden unit value. If the distance is more than 1, represents that the hidden unit's activation value is tending to a different output value in comparison to other clustered hidden unit activation values. Such scattered activation value is not pushed since it contributes to a different output, which should also be considered to maintain the accuracy of the rules.

For example, if there are three hidden units, the activation values for an input sample is 0.245, 0.319 and 0.498. All three values are equidistant to each other and not significantly far to each other. Hence, these values will not be moved closer for accuracy of classification rules. If these values are 0.926, 0.899 and 0.513 then 0.513 will be moved to the calculated distance. In this way all three will contribute to only one rule.

As explained earlier, the error gives the difference between the actual output and the target output for an input sample. If the best performance can be more minimized, the network output will be more closer to target output hence less scattered intermediate values. This will also contribute in reducing the number of rules extracted from the network. The proposed algorithm attempts to work on the same principle to contribute in extracting fewer rules from the network.

3. Data Sets

The proposed method is experimented on five standard datasets from UCI machine learning repository.

ILPD This data set contains 416 liver patient records and 167 non liver patient records. Selector is a class label used to divide into groups(liver patient or not i.e 1 for patient and 0 for non). This data set contains 441 male patient records and 142 female patient records, total 583. A person will have some values for each of these 10 attributes.

WAVEFORM data set consists of 5000 instances of waves. Each wave is characterized by 21 continuous inputs with noise. The problem is to classify these waves into one of three classes.

ARRHYTHMIA This database contains 279 attributes, 206 of which are linear valued and the rest are nominal. The aim is to distinguish between the presence and absence of cardiac arrhythmia and to classify it in one of the 16 groups. automatically processed and the respective diagnostic features measured. The CTGs were also classified by three expert obstetricians and a consensus classification label assigned to each of them. Classification was both with respect to a morphologic pattern (A, B, C. ...) and to a fetal state (N, S, P). Therefore the dataset can be used either for 10-class or 3-class experiments. The instances are divided randomly into three sets: 80 percent for training, 20 percent for testing, and 20 percent for validation.

CTG (CARDIOTOCOGRAPHIC DATA) The dataset consists of measurements of fetal heart rate (FHR) and uterine contraction (UC) features on cardiotocograms classified by expert obstetricians. 2126 fetal cardiotocograms (CTGs) were Here it is considered with 10 classes. The instances are divided randomly into three sets: 80 percent for training, 20 percent for testing, and 20 percent for validation. A three layer feedforward neural network with 11 hidden units is trained on the data. The output of the layer is clustered into 10 groups corresponding to the 10 classes

.

DATA SET	NO.OF. ATTRIBUTESS	NO. OF CLASSES	NO. OF INSTANCES
ILPD	10	2	583
WAVE FORM	21	3	5000
ARRYTHMIA	279	16	452
CTG	21	10	2128
IMAGE SEG	18	7	2310

.

Table 1: Data Sets Used for Evaluation

Image Segmentation: Image data described by high-level numeric valued attributes, 7 classes. The instances were drawn randomly from a database of 7 outdoor images. From each dataset, 80 percent of the data is used for the training the training,10 percent is used for testing and 10 percent for validation.

A three layer feed forward neural network with 8 hidden units is trained on the data. The output of the layer is clustered into 7 groups corresponding to the 7 classes.

4. Experiment and Result

The goal of this evaluation is to compare the number of rules extracted from a trained network when Distance Term is included in the hidden layer (experimental condition) versus the number when Distance Term is not included (control condition). It shows that training with Distance Term produces better separated encoding at the hidden layer, and thus would improve the performance of existing rule extraction methods. The effectiveness of the rule extraction method is evaluated on the data sets having more than 7 classes to classify selected arbitrarily from the UCI Machine Learning Repository . These are large and difficult data sets with many attributes and classes which has not given better results on previous research work[Huynh & Reggia, 2011].

DATA SET	ACTUAL	NEW	%age
ILPD	28.54	15	47.4
WAVE FORM	69.04	39.8	42.4
ARRYTHMIA	32.54839	28.41139	12.7
CTG	165.9231	159.8225	5
IMAGE SEG	84.57	80.68548	5

Table II. Experiment Result of Rule Extraction

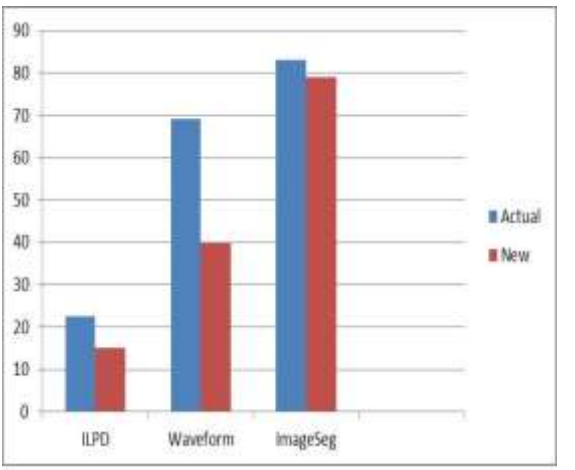

Figure 1

The hidden unit encodings learned by the neural network for five large dataset are used to illustrate proposed method. The proposed method was experimented on datasets for 150 iterations. The results are given below. The result shows that the number of rules extracted are significantly less in number and has not compromised on accuracy.

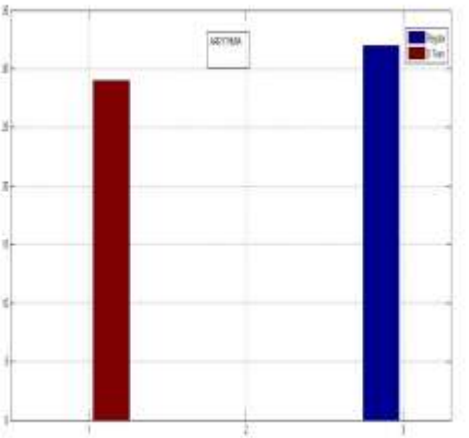

Figure 2

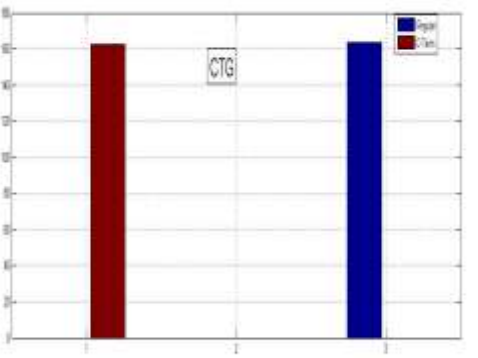

Figure 3

The experiments were repeated multiple times with different initial weights to rule out the effect of randomness and be more confident in the results. Accordingly, the results were averaged over 150 runs in which experiment runs started from the same initial conditions.

Table II shows details of all data sets and the actual number of rules and number of rules with distance term. The less number of rules is helpful to take a decision or conclude about the selector class for a given value of all attributes.

5. Observations and Analysis

Results are analyzed on many parameters namely:

- Number of Classes
  a. Distance moved.
  b. No of neurons moved.
- Number of Instances
- Number of Attributes
- Number of Iterations

S.No	Data Set	Actual	New	Avg Distance Shifted	Total Values Effected
1.	ILPD	28.54	15	1059	453
2.	WAVE FORM	69.04	39.8	49000	2367
3.	ARRYT HMIA	32.54839	28.41 139	4100	1710
4.	CTG	163.9231	159.8 225	96600	160
5.	IMAGE SEG	85.47	80.68 548	45547.5	141

Table III: Experiment Results

Following Observations are made:

- ILPD Dataset has lowest number of attributes, classes and number of instances is also less. The result of ILPD is highest, best among all data sets.
- The waveform has more number of attributes in comparison to ILPD. Number of classes are just one more than ILPD but the number of Instances are almost 100 times more than ILPD and highest among all. This has effected on percentage of reduction in the number of rules.
- Arrythmia has the highest number of attribute and numbers of classes, but instances is lowest even less than ILPD. The percentage of reduction in the number of rules has reduced by more than one third.
- All three parameters of CTG and IMAGE SEGMENTATION are high in numbers. This fact has effected on percentage of reduced the number of rules drastically.
- Below table shows the average distance shifted and total number of neurons shifted towards their resulted class. The value of a neuron will fall in the interval of 0 to 1 for any of the classes of the dataset. Therefore, shifting of the value of a neuron towards its intended class has to be done very judiciously especially when the number of classes is more for a dataset. Waveform, CTG and Image Segmentation dataset have large average distance shifted in comparison to ILPD and Arrythmia but less number of neurons shifted.

6. Comparison of Results:

- ILPD and Waveform have the highest percentage of reduction in the number of rules, whereas CTG and Image Segmentation has the lowest. The difference between them is mainly the number of classes. All parameters of CTG and Image Segmentation are highest.
- Arrhythmia dataset has shown intermediate result. It has the largest number of attributes and classes, but less number of instances.
- ILPD and Waveform dataset has almost the same number of classes for classification. Waveform dataset has more number of attributes and instances in a comparison ILPD dataset, resulted in slight fall in

percentage of reduction in the number of rules.

- CTG and Image Segmentation dataset's all parameters are almost same and hence also has shown the same results.
- Since CTG and Image Segmentation dataset has a high number of attributes as well as instances, it resulted in a higher average distance shifted Since the number of classes is high, it resulted in fewer numbers of shifting neurons.

7. Conclusion

In this paper, five large dataset are used to experiment the performance of the enhanced rule extraction algorithm proposed by [Srivastava *et al.*, 2015]. The results are analyzed to conclude the effect of considered parameters on number of extracted rules. We can summarize the analysis by stating that above results show that the number of attributes and number of instances affects the reduction in the number of rules. The Algorithm works well on a dataset having a lower number of instances and attributes. Similarly classification into the least number of classes gives better results. The number of neurons shifted from one class to another also depends on the number of attributes and instances of the dataset. The number of classes does not affect significantly on it. Hence we can conclude that the reduction in rules depends on the number of values affected of neurons, which are again affected by the number of instances and attributes of the dataset. This analyses may contribute in the application of neural network classification.

8. References Section

A,Gupta. (1999).Generalized Analytic Rule Extraction for Feed forward Neural Networks. *IEEE Transactions on Knowledge and Data Engineering*, 11(6), 985-991.

S.M. Kamruzzaman and A. R. Hasan(2005). Rule Extraction using Artificial Neural Networks. ICTM.

S.M. Kamruzzaman and A.M. Jehad Sarkar(2011) . A new data mining scheme using artificial neural networks Sensors.11, 4622-4647.

T. Q. Huynh and J. A. Reggia (2011). Guiding Hidden Layer Representations for Improved Rule Extraction from Neural Networks, *IEEE, Neuralnetworks*, 22(2), 264-275.

V. Srivastava, C. Dhawale and S. Misra(2015). Enhanced Rule Extraction by Augmenting Internal Representation of Feed Forward Neural Network, *IEEE 978-1-4673-9354-6/15/$31.00 ©2015*

M. M. Mia, S. K. Biswas and M. C. Urmi(2015). An Algorithm For Training Multilayer Perceptron (MLP) For Image Reconstruction Using Neural Network Without Overfitting, IJSTR, 271-275.

A. Asuncion and D.J. Newman, UCI Machine Learning Repository, http:==www.ics.uci.edu/mlearn=MLRepository:html:

L. Ai-sheng, Z.Qi. Automatic modulation classification based on the combination of clustering and neural network, Science Direct, 2011.