International Journal of Computer Applications Technology and Research
Volume 6–Issue 6, 242-244, 2017, ISSN:-2319–8656

Efficacy of OCTAVE Risk Assessment Methodology in Information Systems Organizations

Muhammad Asif Khan
Department of Information Systems
College of Computer Science and Engineering
Taibah University, Madina al Munawwara, Saudi Arabia

Abstract: With the increasing use of computers in business information security has also become a key issue in organizations. Risk assessment in organizations is vital in order to identify threats and take appropriate measures. There are various risk assessment methodologies exist which organizations use for risk assessment depending the type and need of organizations. In this research OCTAVE methodology has been used following a comparative study of various methodologies due to its flexibility and simplicity. The methodology was implemented in a financial institution and results of its efficacy have been discussed.

Keywords: risk; OCTAVE; information systems; security; risk assessment; methodology

1. INTRODUCTION

Information organizations have growing concerns of security of information and associated assets. Now information security is considered the key and prime issues worldwide. Information security is a set of procedures and processes, technology and people which aim to protect assets of organizations [1]. In organizations there are various risks and companies face a major issue that how to evaluate those risks in order to use security controls for removing or mitigating the identified risks [2]. There is no standard methodology or procedure which can be used by organizations. There are numerous risk assessment methods and frameworks [3] and organizations aspiring for security of information need to compare different methodologies and select the best method that suits to their needs. In results of risk assessment organizations measure the severity of risks and develop security controls in order to mitigate the loss and gain maximum benefit from the investment done on security measures. Generally, following identification of risks organizations determine the value of threat, its probability of occurrence and impact the threat may have in organization. The severity of risk can be determined by combining threat occurrence and its impact which can be achieved by applying qualitative, quantitative or both methods at the same time [4].

Organizations may face security threats by various means such as information exposed to hackers on internet, malicious and unscrupulous employees, and breach in physical security. Financial organizations experience financial damage in result of security breach that is sometimes unnoticed due to insignificant security events [5]. As stated earlier, there is no standard methodology or procedure which can be adopted by organizations to determine risk to information security, organizations usually chalk out detailed steps for risk assessment. The proper risk assessment planning helps staff assigned for risk assessment for acting effectively and in a systematic way. In order to assess risks in organizations first risks to the most valuable information assets are prioritized and then level of severity of threat to the assets is evaluated. There are generally two type of risk assessment conducted namely qualitative assessment and quantitative assessment.

A qualitative assessment is the one in which descriptive or relative scale is used to determine probability of occurrence of a threat to an information asset. The evaluators of information assets assess possible threats by drawing some vulnerable scenarios and assign a descriptive scale rather than a numeric value. For example, probability of risk occurrence in a component can be defined as ‘high’, ‘medium’ or ‘low’. This assessment is simple and non-technical people can also be involved in the assessment.

A quantitative assessment uses a numeric value to indicate probability of risk occurrence to an information asset. All risk threat elements are quantified. A probability of occurrence of a risk event is indicated using a numeric value as 35% or 60% etc. Since this method is based on numeric values the calculations may become more complex. This assessment is difficult and employees may find intricacies in understanding it.

2. RISK ASSESSMENT METHODOLOGIES

There are various risk assessment methodologies used by different organizations depending on the type and need of organizations. Some methodologies require large enterprises and experienced staff to use them as they are quite sophisticated to be used. Most of the methodologies are commercially developed, therefore, unavailable to public except some for marketing purpose. An organization needs to use a methodology for risk assessment usually intends to compare different methodologies before the right one cold be selected. But investment on purchasing different methodologies for comparison purpose is not viable. Since documentation and presentations on different methodologies are available comparison, most of the time comparison is made through such material. Table 1 shows a list of risk assessment methodologies that are commonly used in organizations

Table 1. Risk assessment methodologies

Methodology	Description
Asset Audit	In order to determine whether assets of a company have a potential threat. It also determines likelihood of occurrence of a threat and impact of threat
CORAS	A qualitative model-based methodology consists of four diagrams – an extension of UML. It requires expert knowledge to use for risk assessment
CRAMM	This is qualitative methodology that focuses on assets and valuation. Following valuation of assets likelihood of threat is determined. It requires experts to use
OCTAVE	A qualitative simple methodology that can be used by knowledgeable small team of business and IT people. It is not driven by technology but practices of security and risk which propagate main information of security
NIST	A qualitative or quantitative methodology that is cost effective and quick in assessing security within organizations via survey instruments
Risk IT	A complementary part of COBIT framework developed by ISACA. It provides guidelines for IT security as well as risk assessment which cannot be used freely without using the framework

It can be observed that all risk assessment methodologies have some disparities in terms of scope or application. Some of them require expertise and thorough knowledge whereas some methodologies work in combination of generic framework of risk assessment.

2.1 OCTAVE methodology

Operationally Critical Threat, Asset, Vulnerability and Evaluation (OCTAVE) methodology was develop at Carnegie Mellon University, USA [6]. This methodology is used in small to medium organizations. It can be tailored according to an organization environment. Using the methodology firms can reduce overhead cost spent on training and knowledge development that are required for risk assessment. This methodology is comprised of three phases and each phase consists of a number of processes.

2.1.1 Phase 1 - Build asset profiles

In this first phase all the important assets, prevailing security practices and vulnerabilities in organizations are identified. Also staff knowledge about assets, their vulnerabilities and current security strategies are identified. Based on such information most important vulnerable assets are sorted out.

2.1.2 Phase 2– Identify infrastructure vulnerabilities

In the second phase infrastructure of the organization is evaluated in order to determine technological vulnerabilities which could harm important assets. The components which are most critical are further evaluated and technological weaknesses are detected.

2.1.3 Phase 3 – Develop security strategy and plans

In the third phase security risks in the assets identified in the previous phase are mitigated if not removed completely. In order to evaluate impact of threats to the assets criteria are developed which in turn gives risk profile to each asset. Finally, a strategy for protection of assets is developed and an approval is requested from the management.

3. EFFICACY OF OCTAVE METHOD

As discussed earlier OCTAVE methodology is a simple that can be used with small team of knowledgeable employees within an organization. In order to determine the risk assessment in a financial organization in Saudi Arabia this method is used and for this purpose all three phases of the methodology were implemented step by step. To start with the OCTAVE methodology initially two teams were formed within the financial institution i.e. one from the business department and another from IT department. In both the teams members were knowledgeable in their respective fields and capable of giving answers to the questions related to security and vulnerability. To start with the research study different sessions, discussions and interviews were conducted with both the teams in order to know vulnerable assets and current strategy to protect them. After collecting information critical assets with vulnerability were evaluated and further detailed discussions were held with the management in order to ensure criticality of the assets. IT team provided sufficient information in terms of technologies that may have caused the assets vulnerable. As the last phase of the methodology suggests for developing a security strategy to protect the critical assets, a security strategic plan was developed and presented to the top management for review and approval. The OCTAVE methodology provided clear and transparent guidelines to evaluate vulnerable assets in the financial institution and helped to build a viable and useful security plan for the critical assets. The methodology was successfully implemented and Figure 1 shows the process

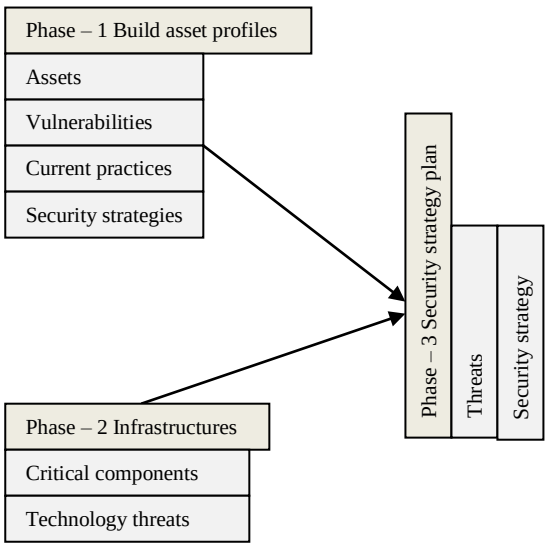


Figure. 1 Implemented OCTAVE methodology

4. RESULTS AND DISCUSSION

Since OCTAVE is a qualitative methodology a relative scale is used for determining probability of a risk to be occurred. Table II shows the scale used in determining the occurrence of risk in critical assets

Table II. Scale to determine risk probability

Probability	Description
Very high	Threat has occurred earlier and it is likely to occur in the present condition
High	Threat has occurred in the past and it is likely to occur
Normal	Threat may occur
Low	Threat occurred seldom in the past and most likely not to occur
Very low	Threat very unlikely to occur and may occur in unusual circumstance

When the staff members of the financial institution discussed different critical assets they were provided with the above scale to gauge risk in the assets. Table III shows the data obtained by the staff about the risk assets

Table 2II. Data about risk probability

Threat	Probability	Description
Hacker	Low	Intruder may access data or deny accessibility to data
Theft	Very low	A person/employee may steal data or devices physically
Data integrity	Very low	Information can be altered without authorization
Authorization	High	Accessibility to physical system without permission
Firewalls	Very low	Insufficient security to protect systems and data
Virus/worm	Low	Spread of malicious programs within organization
Disaster	Very low	An insider or outsider of the organization may destroy data
Cloud computing	Normal	Concerns of data security over cloud
Encryption	Very low	Encoded data captured and modified
Denial of service	Normal	Unavailability of service to legitimate customers or users

The above data clearly shows that assets in the financial institutions are somehow secure, although some threats need to be handled properly. For example, during discussion it was informed that sometime unauthorized personnel enter in operational areas for socializing with friends which may be threat for information assets. This methodology provided an opportunity to the management for making strict policy to ban entry to unauthorized people in operational areas. Similarly, latest technologies and updated versions of software are needed to secure the assets as at times service was unavailable to customers for some time and hackers blocked the traffic of data to pass through the servers of the institution. By this methodology the management was able to identify various vulnerable assets and obsolete technologies to be updated or replaced. At the end management was recommended a to prepare a detailed security strategy to protect the information assets.

In future, researchers may explore other type of organizations and threats and assess with different methodology. The methodology used in this study quite simple and easy to implement and can be used further in different type of organizations.

5. REFERENCES

[1] Jourdan, Z, Rainer, K., Marshall, E., and Ford, N. 2010 An investigation of organizational information security risk analysis. Journal of Service Science. 3, 33-42

[2] Syalim, A., Hori, Y., and Sakurai, K. 2009 Comparison of risk analysis methods: Microsoft' security management guide. International conference on availability, reliability and security. 726-731

[3] Saleh, S., and Alfantookh, A. 2011 A new comprehensive framework for enterprise information security risk management. Applied Computing and Informatics. 9, 107-118

[4] Palaniappan, S., Rabiah, A., and Mariana, Y. 2013 A conceptual framework of info structure for information security risk assessment. Journal of Information Security and Applications. 18, 45-52

[5] Ben, R., Jouini, M., Ben, A., and Milli, A. 2012 A cyber security model in cloud computing environments. Journal of King Saud University. 1, 63-75

[6] OCTAVE, http://www.cert.org/resilience/products-services/octave/index.cfm", [Retrieved on May 23, 2017]

International Journal of Computer Applications Technology and Research
Volume 6–Issue 6, 245-248, 2017, ISSN:-2319–8656

Object tracking with SURF: ARM-Based platform Implementation

H. Hassnaoui
EEA&TI laboratory, Hassan II University of Casablanca Faculty of Sciences and Techniques (FSTM)
Mohammedia, Morocco

A. Badri
EEA&TI laboratory, Hassan II University of Casablanca Faculty of Sciences and Techniques (FSTM)
Mohammedia, Morocco

A. Sahel
EEA&TI laboratory, Hassan II University of Casablanca Faculty of Sciences and Techniques (FSTM)
Mohammedia, Morocco

M. Akil
2ESIEE-Paris, 93162
Noisy-le-Grand Cedex, France

Abstract: Several algorithms for object tracking, are developed, but our method is slightly different, it's about how to adapt and implement such algorithms on mobile platform.
We started our work by studying and analyzing feature matching algorithms, to highlight the most appropriate implementation technique for our case.
In this paper, we propose a technique of implementation of the algorithm SURF (Speeded Up Robust Features), for purposes of recognition and object tracking in real time. This is achieved by the realization of an application on a mobile platform such a Raspberry pi, when we can select an image containing the object to be tracked, in the scene captured by the live camera pi. Our algorithm calculates the SURF descriptor for the two images to detect the similarity therebetween, and then matching between similar objects. In the second level, we extend our algorithm to achieve a tracking in real time, all that must respect raspberry pi performances. So, the first thing is setting up all libraries that the raspberry pi need, then adapt the algorithm with card's performances. This paper presents experimental results on a set of evaluation images as well as images obtained in real time.

Keywords: object tracking, mobile platform, feature matching, SURF, Raspberry pi

1. INTRODUCTION

In the field of computer vision, we seek the improvement of perception and visual recognition, by studying the algorithms carried out in this sense, and thus propose an adequate technique of implementation. Several techniques have been discussed to improve artificial vision. The different methods are based on image's content analysis, to extract the interest areas for studying. The descriptors of its zones are calculated independently on the scale (scale invariant) and rotation, to have the necessary information, which will be exploited to compare the images. That way we can detect and track objects. In the literature, the first method introducing the notion of independence at scale and rotation is that proposed by researcher David Lowe in 1999, called SIFT (Scale Invariant Features Transform). It is about detecting zones in an image which are known as interest points by DOG (Difference Of Gaussians) method and then, for each point, a descriptor vector of 128 dimensions is computed which is set The relation of this pixel with its neighborhood in the different scales or resolutions. This method is robust but has a major disadvantage that resides in the calculation time which is important which influences detection in real time. To remedy this problem, in 2006 researchers "ETH Zurich and Katholieke Universiteit Leuven" proposed an accelerated technique inspired by the SIFT which named SURF (Speeded Up Robust Features). SURF is based on Haar 2D wavelet responses and uses the integral images. As a basic characteristic, SURF uses a Haar wavelet approximation of blob detector based on the determinant of the Hessian matrix. These two methods are used for object detection or 3D reconstruction. Our work consists in proposing a method of implementing these techniques in a mobile platform while improving the calculation time to adapt the algorithm to such a platform. First, we will present the two methods (SIFT and SURF), after we will propose our technique of implementation in a mobile platform respecting the software architecture of the platform and its physical performances.

2. RELATED RESEARCHES

The first work on objects recognition is begun by the interest points extraction. An interest point is a point where the contour direction of an object changes abruptly (corner) or an intersection between two (or more) contour segments.

Moravec [1] considered a window as a neighborhood of the pixel, then determine the mean changes of the intensity in the neighborhood considered, moving the window in various directions. One of the main problems with this operator is that is not isotropic: If an edge is present that is not in the direction of the neighbors, then the edge will be badly chosen as an interest point. So, the Moravec operator is sensitive to noise. To remedy the problem of anisotropy Harris [2] proposes using a Gaussian filter W rather than a binary filter (0 or 1) used by Moravec. By these two methods it is possible to detect only the objects which have the same resolution and same angle of rotation. To solve this problem D. LOWE [5] proposes a robust algorithm called SIFT (Scale Invariant Features Transform) that is invariant with scale and rotation. This technique has the disadvantage of computing time which is important hence the invention of a new SURF [11] method inspired by the SIFT but it is three times faster.

3. FEATURE MATCHING

3.1 SIFT

The first step of the algorithm is interest points detection, called key points. A key point (x, y, σ) is defined by its coordinates on the image (x, y) and by its characteristic scale factor (σ). This is a circular interest zone; which radius is proportional to the scale factor. Each key point associated with an intrinsic orientation, dependent on the local content of the image around the key point, with the scale factor considered.

Detection and retrieval of features of interest points are carried out in four steps:

• Detection of extrema of scale-space

• Location of interest points,

• Choice of descriptor orientation,

• Calculation of descriptors

3.2 SURF

SURF detector is based on the Hessian matrix because of its good performance in computation time and accuracy [11]. Given a point x = (x, y) in an image I, the Hessian matrix H(x,σ) in x at scale σ is defined as follows:

$$H(x,\sigma) = \begin{pmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{pmatrix}$$

Where Lxx(x, σ) is the convolution of the Gaussian second order derivative ∂2/∂x2 g(σ) with the image I in point x, and similarly for Lxy(x, σ) and Lyy(x, σ). In practice, however, the Gaussian needs to be discretized and finite. On the next image you can see the partial derivatives of the Gaussian. First finite and discretized (the two left images) and then approximated by a 'box filter' in the directions y and xy. The gray areas are zero.

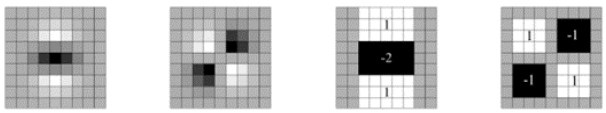


Figure 1: partial second derivatives of the Gaussian [11]

The approximation of the determinant of the Hessian matrix calculated in a point x of the image is stored in a “blob response map” and then local maxima are searched to find blobs.

It is interesting to find different scales to interest points to make the detector invariant to scale changes. This is often taken into account by creating a pyramid of images.

Each level of the pyramid represents a different scale. SURF can proceed differently through the box filters. Instead of successively applying the same filter to the output image filtered and sub-sampled, we can use box filters of various sizes directly on the original image. The “blob response maps” to different scales are constructed by enlarging the filter rather than repeatedly reducing the image size. This allows one hand to reduce the computation time and also prevent aliasing due to under - sampling of the image. The left image of the figure below shows the classical method with sub-sampling filter and constant size. On the right image filters vary in size.

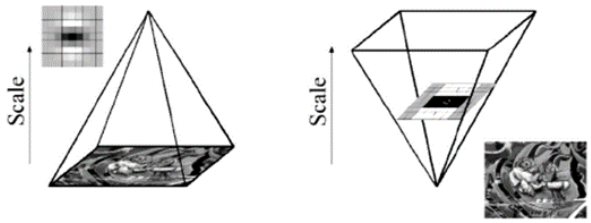


Figure 2: multiscale detection [11]

Scale spaces in SURF are implemented by applying box filters of different sizes. The output of the above 9×9 filter is considered as the initial scale layer, to which we will refer as scale s=1.2 (corresponding to Gaussian derivatives with σ=1.2). Specifically, this results in filters of size 9×9, 15×15, 21×21, 27×27, etc.

In searching for the maxima of the “blob response map” at different levels; we can now extract the position and size of the blobs in the image.

Local neighborhood descriptor:

The goal of a descriptor is to provide a unique and robust description of an image feature by describing the intensity distribution of the pixels within the neighbourhood of the interest point. A description is obtained for every interest point identified previously.

Orientation:

To achieve rotational invariance, the orientation of the interest point needs to be found. Bay et al. used The Haar wavelet responses in both x and y directions within a circular neighborhood of radius 6s around the interest point are computed, where s is the scale at which the interest point was detected. The dominant orientation is estimated by calculating the sum of all responses within a sliding orientation window of size π/3. The horizontal and vertical responses within the window are summed. The two summed responses then yield a local orientation vector. The longest such vector overall defines the orientation of the interest point.

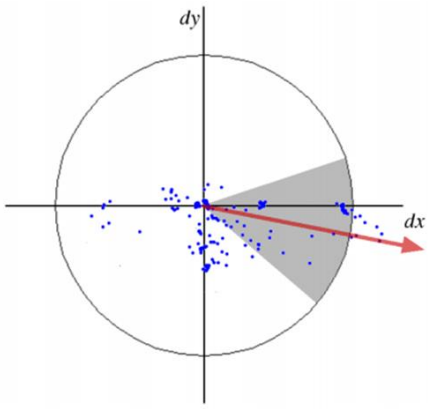


Figure. 3: Orientation assignment: A sliding orientation window of size π/3 detects the dominant orientation of the Gaussian weighted Haar wavelet responses at every sample point within a circular neighbourhood around the interest point. [11]

By comparing descriptors obtained from different images, *matching* pairs can be found.

4. IMPLEMENTATION

4.1 Raspberry pi

The Raspberry Pi is a single-board nano-computer with ARM processor, it allows the execution of several variants of the free operating system GNU / Linux and compatible software. In 2006, the first prototypes of the Raspberry Pi were developed on Atmega 644 ATmel microcontrollers. Different versions appeared: Model A, A +, B, B +, PI 2 and PI 3. The following table presents some technical characteristics of the pi version 3.

Table 1. Technical characteristics of Raspberry Pi 3

technical characteristics of Raspberry Pi 3	
Input/output interface	**Hardware Performance**
• 4 2.0 USB ports • 10/100 Mb Ethernet • HDMI • Audio Outputs (3.5 mm phone jack) • Storage: Micro SDHC slot • Power supply: micro USB • 40 I/O GPIO (General Purpose Input Output) with I2C, SPI, S2C, PWM…	• CPU: 1.2 GHz 64/32-bit quad-core ARM Cortex-A53 • 1 Go RAM • GPU BCM videocore 4 full HD 1080p 30 fps

4.2 Application

Our contribution is to adapt the matching algorithms to the Rasperry pi in real time. First, we looked for a suitable programming method for the card, installing all necessary libraries, we used the camera pi 8 Mp to acquire the images. Our objective is to find similar objects in the different images, so first, we must load or capture the image, containing the object to be detected or to follow, on which we apply the SURF algorithm to extract the interest points and then calculate the descriptors for these points. Then the same algorithm (SURF) will be applied on a scene image which we want to examine to know if the object in question exists or not. Finally, we apply the matching algorithm between the similar descriptors and thus between the similar areas in the two images (FIG. 4). To extend our application in real time we add a small improvement. Indeed, we define a video capture loop with a rate of 15 fps witch the SURF algorithm is applied to each of these frames (FIG. 5).

Our problem lies in the real-time operation which causes an overflow problem, because we overcome the performance of the card, to remedy it we deactivated the rotation parameter of SURF. Therefore, the algorithm is deprived now of the invariance to the rotation.

5. FIGURES

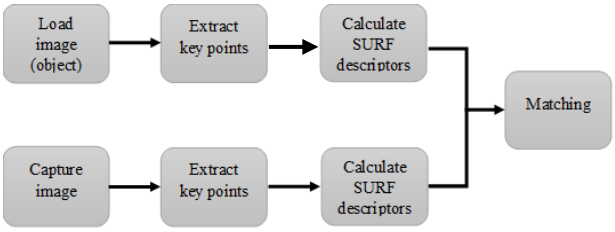


Figure. 4: object detection algorithm: captured image

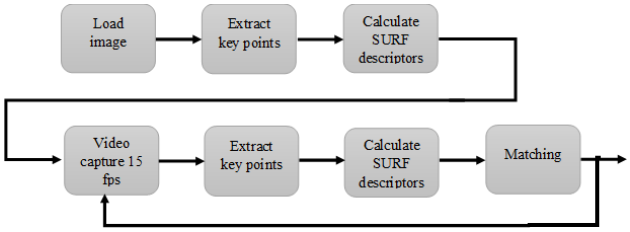


Figure. 5: object detection algorithm in real time

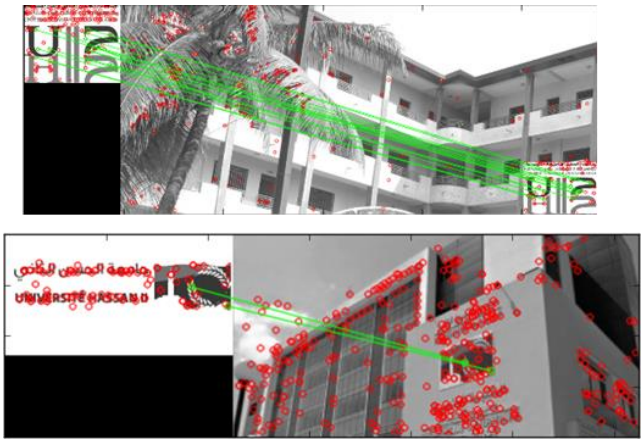


Figure. 6: object detection and matching using Pi camera and Raspberry pi 3

6. CONCLUSION

We have presented a technique of implementation of SURF algorithm in Raspberry pi, in order to detect objects in real time, with CPU computing on ARM-Based platform (Raspberry pi). Our algorithm brings together efficiency, speed and also portability. As such it is possible to optimize the artificial vision in the industrial sector.

This application has given a satisfactory results. In future work we will aim at optimizing the algorithm by parallel computing on GPU in order to combine SURF with HDR imaging in such platform.

7. REFERENCES

[1] Moravec, H. 1980. Obstacle avoidance and navigation in the real world by a seeing robot rover. In tech. report CMU-RI-TR-80-03, Robotics Institute, Carnegie Mellon University and doctoral dissertation, Stanford University, number CMU-RI-TR-80-03.

[2] Harris, C. and Stephens, M. 1988. A combined corner and edge detector. In Proc. Fourth Alvey Vision Conference, pages 147–151.

[3] Lindeberg, T. 1988. Feature detection with automatic scale selection. IJCV 30(2).

[4] Mikolajczyk, K. 2002. Interest point detection invariant to affine transformations. PhD thesis,Institut National Polytechnique de Grenoble.

[5] Lowe, D.G. 1999. Object recognition from local scale-invariant features. In Proceedings of the 7th International Conference on Computer Vision, Kerkyra, Greece, pp. 1150–1157.

[6] David, G. Lowe, 2004. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, vol. 60, no 2, 2004, p. 91–110.

[7] Brown, M. and Lowe, D.G. 2002. Invariant features from interest point groups. In British Machine Vision Conference, BMVC 2002, Cardiff, Wales, pp. 656-665.

[8] Lindeberg, T. and Bretzner, L. 2003. Real-time scale selection in hybrid multi-scale representations. In Proceedings of the 4th international conference on Scale space methods in computer vision, vol. 2695, Berlin, Springer-Verlag.

[9] Lowe, D.G. 2001. Local feature view clustering for 3D object recognition. IEEE Conference on ComputerVision and Pattern Recognition, Kauai, Hawaii, pp. 682-688.

[10] Ke, Y. and Sukthankar, R. 2004. PCA-SIFT: A More Distinctive Representation for Local Image Descriptors. Proc. Conf. Computer Vision and Pattern Recognition, pp. 511-517.

[11] Bay, H., Ess, A., Tuytelaars, T. and Van Gool, L. 2008. SURF: Speeded Up Robust Features. Computer Vision and Image Understanding (CVIU), Vol. 110, No. 3, pp. 346—359.

[12] Fan, X. et al. 2013. Implementation of high performance Hardware architecture of OpenSURF Algorithm on FPGA. International Conference on Field-Programmable Technology (FPT). IEEE Conference Publications.

[13] Wilson, C et al. 2014. A Power-Efficient Real-Time Architecture for SURF Feature Extraction. International Conference on ReConFigurable Computing and FPGAs (ReConFig14). IEEE Conference Publications.

[14] Chao, J., Huitl, R., Steinbach, E. and Schroeder, D. 2015. A novel rate control framework for sift/surf feature preservation in h.264/avc video compression. IEEE Transactions on Circuits and Systems for Video Technology, vol. 25, no. 6, pp. 958–972.

[15] Chao, J et al. 2015. A Novel Rate Control Framework for SIFT/SURF Feature Preservation in H.264/AVC Video Compression. IEEE Transactions On Circuits And Systems For Video Technology, Vol. 25, no. 6.

[16] Chen, W. et al. 2016. FPGA-Based Parallel Implementation of SURF Algorithm. IEEE 22nd International Conference on Parallel and Distributed Systems (ICPADS) Pages: 308 – 315.

International Journal of Computer Applications Technology and Research
Volume 6–Issue 6, 249-253, 2017, ISSN:-2319–8656

Adding Remote Controller Functionality To Any Stereo

Joe Abraham
Department of Electronics andCommunication Engineering
Jyothi Engineering College, Cheruthuruthy
Thrissur-679531, Kerala, India

Ambily Francis
Department of Electronics andCommunication Engineering
Jyothi Engineering College, Cheruthuruthy
Thrissur-679531, Kerala, India

Abstract:. Use of stereo has become common in our lives. They are used in cars, TVs, music players etc. And it is essential at least to control their volumes. Suppose there is a stereo amplifier which functions pretty well but it does not have a remote. It would be very annoying if its volume cannot be controlled. So this project is useful as it creates a device which makes use of any existing remote to control the volume. For controlling the volume, we use a volume controller IC. The electronic volume controller IC PT2258 is a digital potentiometer which can be controlled using I2C protocols. It is used to control the attenuation for every combination possible from 0 to -79 dB/step. Universal IR receiver is used to decode the IR codes and the data will be transferred to the Arduino which in turn communicates with the IC PT2258 and controls the volume. The device also consists of two buttons, which are used to synchronize the IR code of the existing remote with the device. So the user will be able to use the device easily.

Keywords:

1. INTRODUCTION

The device designed is used to control the stereo which doesn't come up with a remote controller or a device which is so old that the remote controller is not available in the market. It will help the users to control the volume of any stereo with any remote they have got with them. So to be precise we can add up a remote controllable system to a normal computer speaker or any speaker that are readily available in the market which doesn't come with remote and we need to control the volume of the same using a remote controller.

The device makes use of an Arduino, IC PT2258 which is a digital potentiometer, Universal IR receiver, two switches and 4 audio jacks (2 for input and 2 for output). The Universal IR receiver receives the code and is given to the Arduino. Arduino takes the value and controls the IC PT2258.The volume is controlled by making attenuation in the channels. Thus the output of the device will be attenuated signal of the input.

The same device can also be made using a pulse detector and an analog potentiometer IC, but the problem is that, we need to provide a predefined remote controller along with the device. Else we have to redesign the pulse detector each time the user needs to use any different remotecontrollers. On using an Arduino, the predefined IR codes can be changed simply by pressing the switch in the device. By pressing the switch new IR values will be written to the device.

The main application of this device is that in a stereo, which does not come with any remote or whose remote is lost or damaged but they work fine manually, can be controlled with any available remote

2. LITERATURE SURVEY

2.1 Humble Volume Control Circuit

The humble volume control circuit can use to control the volume. Volume controls are usually implemented with a potentiometer configured as a voltage divider. The signal goes to lug 1, the output is connected to the wiper (lug 2) and the lug 3 is connected to ground. If we turn the pot shaft, more or less of the signal voltage is sent to either to ground (quieter) or to the output through the wiper (louder). This simple approach works well, but we are concerned with stereo, which has that pesky concept of two independent channels. But your standard pot only is a "single-gang" device: it only has one set of lugs.

There are many issues with dual gang pots. Potentiometers are not exact devices. Their tolerance ratings aren't terribly high, usually in the range of 10percentage−20percentage. That means that you can grab two 10K ohm pots of the exact same part and manufacturer and expect one pot to max out at 8k and the other at 12k. This really isn^{0}t a big deal for most applications circuits are designed with part tolerances in mind.

Stereo input volume control with those variances. Each "gang" of the pot could be off by as much as 20 percentages, and matching of both gangs with a given shaft position could be way off also, Make sense in an empirical way. The result is that by using a cheap dual-gang pot as a volume control you are actually building in an unintentional balance control also. You can never really be sure if the left and right levels being reproduced are what they were when recorded. The problem is these variances can become quite pronounced when you are dealing a stereo volume control. But in our project attenuation is properly controlled by IC PT228 by different attenuation steps so that balance control problem is minor

2.2 Balanced Volume Control Circuit

The volume balance control circuit is a circuit that using to control the music between its source and speakers. The circuit is based on LM1036N which is a DC controlled tone (bass/treble), volume and balance circuit for stereo applications in car radio, TV and audio systems. The main disadvantages are complexity in design and are expensive.

2.3 Digital Volume Control Circuit

Digital volume control circuit is the another popular circuit using the IC MAX5486.

MAX5486 is a 40K dual digital volume / balance controller that has a pushbutton interface. The IC has a built in bias voltage source that eliminated the need of an external circuitry for the same purpose and thereby by reduces external parts count. The IC also has an LED status indicator driver circuit which can be used for driving the status indicator LEDs which indicates the volume level and balance level. The IC can be operated from a single or dual power supply and is available in 24 pin TSSOP package. The volume control circuit based on MAX5486 can be applied in a lot application like personal audio systems, hand held audio devices, home theatre systems, car audio systems,**computer audio systems etc. The IC is not really purposeful an accuracy will be less But the main advantages of our project is, we can control the volume, without the circuit in hand, using a remote. The synchronized working of both IR sensors controls volume with any remote. It is an easy and advanced method comparing with other old methods.**

3 CIRCUIT DIAGRAM

3.1 OPERATON

There are two switches in the device. By pressing the first switch and by pressing any key in the remote controller will make IR code emitted by remote save as Volume up command. And by pressing the second switch and then pressing any switch in remote will make the IR code save as volume down command. Once the value is saved the device will be comparing these values with the received IR values. If an IR code is received is same as that of a volume up command, then device will decrease the attenuation. And if the IR code received is same as that of volume down command then the device will increase the attenuation. The universal IR receiver receives the IR code and is given to the Arduino. The Arduino compares the received values to the predetermined values. The Arduino accepts the IR code if it is a predetermined one else it ignores the IR code and receives the next IR code. If the IR code is predetermined one, then the Arduino compares it with volume up command and volume down command. An attenuation factor is defined in the program. If a volume up command is received, then the attenuation factor will be decreased and if a volume down command is received then the attenuation factor will be increased. The value of attenuation factor is limited between 0 to 79 db. Once attenuation factor has been changed it will be communicated to the IC PT2258 using SDA and SCL which are used for I2C communication. For this wire library is used in Arduino. Attenuation factor is changed means that an attenuation is given in the output signal when compared to the input. This simply means the volume has been changed or decreased according to the attenuation factor. This is how the device controls.

IC PT2258

PT2258 is a 6 channel electric volume controller IC utilizing CMOS technology, specially designed for the new generation of AV multi-channel Audio system. The attenuation ranges 0 to 79dB at 1dB step, low noise high channel separation.

BUS INTERFACE

Data are transmitted to and from the microprocessor to the PT2258 via the SDL and SCL make up the bus interface. It should be noted that the pull up resistors must be connected to the positive supply voltage.

DATA VALIDITY

The Data on the SDA line is consider valid and stable only when the SCL signal is in high state the high and low state SDA line can only change when SCL signal is low.

Fig 2 : Internal Block Diagram of PT2258

START CONDITIONS

A Start Condition is activated when,

The SCL is set to HIGH and SDA shifts from HIGH to LOW State.

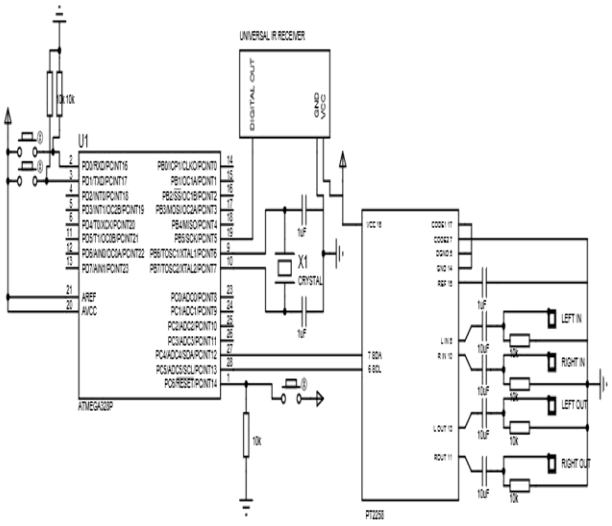


Figure 1: Diagram Representing the Whole of the Circuit Diagram

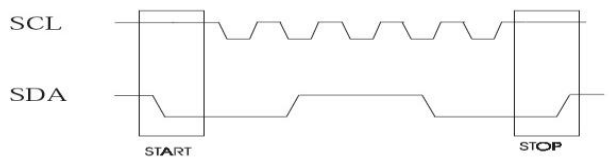


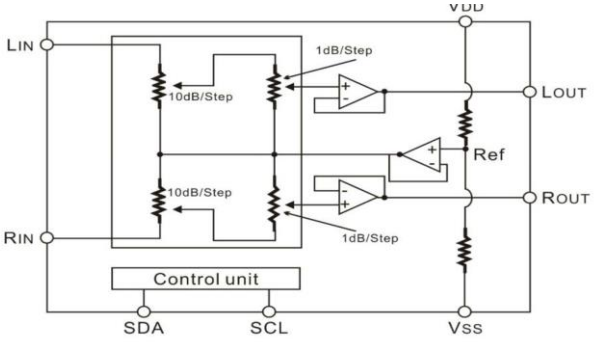
Figure 3: Start and Stop Condition Diagram

STOP CONDITION

The Stop Condition is activated when,SCL is set to HIGH and SDA shifts from LOW to HIGH State.Every byte transmitted to the SDA Line consists of 8 bits.

SOFTWARE DESCRIPTION

PT2258 Address Code depends on the state of CODE1 (Pin No. 17) and CODE2 (Pin No.4). If CODE1 or CODE2 is connected to Vcc, then CODE1 or CODE2 is set



to 1. If CODE1 or CODE2 is connected to the Ground, it is set to 0. Data byte Description.Every byte transmitted to SDA line consist of 8 bites. Each byte must be followed by acknowledge bit. The MSB is transmitted first.

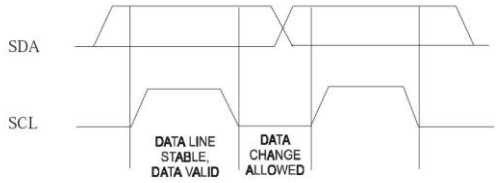


Figure 3: Start and Stop Condition Diagram

FUNCTIONAL DESCRIPTION

Data are transmitted to and from the microprocessor to the PT2258 via the SDA and SCL.The SDA and SCL make up the BUS Interface.A data on the SDA Line is considered valid and stable only when the SCL Signal is in HIGH State. The HIGH and LOW States of the SDA Line can only change when the SCL signal is LOW.

INTERFACE PROTOCOL

- A start Condition
- A chip Address Byte including PT2258 address.8th bit of the byte must be 0. It must always acknowledge the end transmitted byte.
- Data sequence (N-bytes+Acknowledge)

o IR Sensor

IR used in this device is a universal IR receiver. Universal IR receiver is used to read the infra-red code. It converts the infra-red code to digital signal a transmits it through the digital pin to the micro-controller or microprocessor.

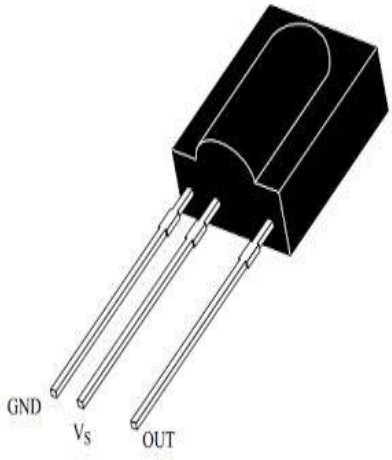


Figure 6: Pin Out of IR Sensor

MSB							LSB	
1	1	1	0	A3	A2	A1	A0	6-Channel- -1dB/step
1	1	0	1	0	B2	B1	B0	6- channel- -10dB/step
0	0	0	1	A3	A2	A1	A0	channel No.3- -1dB/step
0	0	0	0	0	B2	B1	B0	channel No.3- -10dB/step
0	0	1	1	A3	A2	A1	A0	channel no.4- -1dB/step
0	0	1	0	0	B2	B1	B0	channel no.4- -10dB/step
0	1	0	1	A3	A2	A1	A0	channel No.2- -1dB/step
0	1	0	0	0	B2	B1	B0	channel No.2- -10dB/step
0	1	1	1	A3	A2	A1	A0	channel No.5- -1dB/step
0	1	1	0	0	B2	B1	B0	channel No.5- -10dB/decade
1	0	0	1	A3	A2	A1	A0	channel No.1- -1dB/step
1	0	0	0	0	B2	B1	B0	channel No.1- -10dB/step
1	0	1	1	A3	A2	A1	A0	channel No.6- -1dB/step
1	0	1	0	0	B2	B1	B0	channel No.6- -10dB/step
1	1	1	1	1	0	0	M	6-Channel, M=1 MUTE ON, M=0 MUTE OFF

ATTENUATION UNIT BIT

A3/...	A2/B2	A1/B1	A0/B0	ATTENUATION VALUE(dB)+
0	0	0	0	0/0
0	0	0	1	-1/-10
0	0	1	0	-2/-20
0	0	1	1	-3/-30
0	1	0	0	-4/-40
0	1	0	1	-5/-50
0	1	1	0	-6/-60
0	1	1	1	-7/-70
1	0	0	0	-8/...
1	0	0	1	-9/...

Arduino and Programming

Algorithm

1. Start
2. Check the buttonstate1 and buttonstate2.If any one of them is high, save the IR code received as a and b respectively
3. Read the IR codes available at the IR sensor
4. Compare it with a and b.
5. If IR code received is a.
6. And if the attenuation factor is i is greater than 0,Then i = i -5,and jump to step 10
7. If IR code received is b
8. And if the attenuation factor is i is lesser than 79,Then i = i +5,and jump to step 10
9. If the IR code is not a and b, then ignore them.
10. Split the attenuation factor i to ones and tens.
11. Transmit this ones and tens to the PT2258 IC through I2C channels.
12. Jump to step2

Hardware Description

IC PT2258

PT2258 is a 6-Channel Electronic Volume Controller IC utilizing CMOS Technology specially designed for the new generation of AV Multi-Channel Audio System. PT2258 provides an I2 C Control Interface, an attenuation range of 0 to -79 dB at 1 dB/step, low noise, high channel separation. Housed in 20-pin, DIP or SO Package, PT2258's pin assignments and´ application circuit are optimized for easy PCB Layout and cost saving advantages.

The main features of IC PT2258 are as follows:

- CMOS Technology
- Low Power Consumption
- Least External Components
- Attenuation Range: 0 to -79 dB at 1dB/step
- Operating Voltage: 5 to 9v Low Noise, S/N Ratio>100dB (A-weighting)
- High Channel Separation
- I2 C Bus Control Interface
- Selectable Address
- 6-Channel Outputs

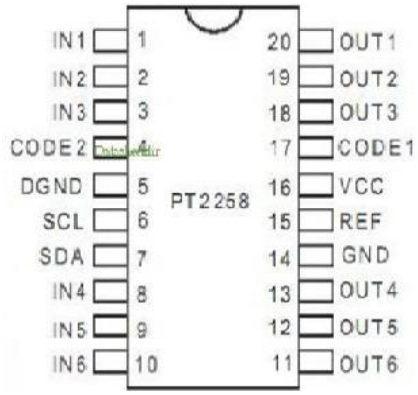


Figure 5.1: Pinout Diagram Of IC PT2258

IC PT2258 has any applications such as:

- AV Surround Audio Equipment
- Car Audio
- Mini Components
- Computer Multi-Media Speaker
- Other Audio Equipments

IR Sensor

An Infra red (IR) sensor is used to detect obstacles in front of the robot or to differentiate between colors depending on the configuration of the sensor. The picture shown is a very simple black box model of the IR Sensor. The sensor emits IR light and gives a signal when it detects the reflected light.It is a universal IR receiver which receivers all the IR codes that comes to it and it consist of 3 pins.It is a universal IR receiver which receivers all the IR codes that comes to it and it consist of 3 pins. Vcc, GND and digital out. Digital output gives the IR code received digitally.This is a simple infra red receiver which receives all the IR codes.It doesnt come specific.It is a universal on. It receives the IR code and transmit the received data digitally through out the digital pin.

Figure 5.2: Ir Sensor Block Diagram

RESULT

The device was designed and implemented. And the device works fine with almost all remote controllers available in the market.

4. CONCLUSION AND FUTURESCOPE

A device is designed and implemented in such a way that it can use any existing remote, to control the volume of a remote less stereo or can be used instead of a damaged remote. In this project we have studied how to control the volume using volume control IC, PT2258. Basic understanding about the IR sensor and Arduino was also studied during this work. The design and verification of the remote control device was successful. The main advantage of this device is that any available remote can be used to control the volume of the required stereo just by pressing the buttons on the device implemented. Thus it is useful to operate even an old, good functioning stereo as it helps to control the volume. The project gave practical knowledge about the electronic circuit and components used in this work and helped to develop new ideas with proper guidance and help of facilities from the department. Project got completed successfully. Circuit can have further implementation in the future.

5. REFERENCES

[1] G. Eason, B. Noble, and I.N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529-551, April 1955. (*references*)

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68-73.

[3] I.S. Jacobs and C.P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

[4] K. Elissa, “Title of paper if known,” unpublished

[5] R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[6] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

International Journal of Computer Applications Technology and Research
Volume 6–Issue 6, 254-258, 2017, ISSN:-2319–8656

Audio Steganography Using Tone Insertion Technique

Suhaip A. Yousif
Karrary University
Khartoum, Sudan

Talaat M.wahbi
Sudan University of
Science and Technology
Khartoum, Sudan

Mohamed H. Sayed
The National Ribat
University
Khartoum, Sudan

Abstract: This paper presents a new technique of embedding text data into an audio file using tone insertion method. The new technique generates two frequency f1 and f2 and inserts them into audio file in a suitable power level according to specific table (stego-table). The proposed technique aimed to increase the payload capacity of the audio file using two bits in the frame without increases the number of the inserted frequencies to four frequencies, as well as using another convoy frequency (CF) for specific pattern. The proposed method conceals the English text into the .wave audio. The performance of the proposed method has been checked by spectrogram, MSE and PSNR.

Keywords: Audio steganography, Tone insertion, spectrogram, PSNR, MSE.

1. INTRODUCTION

Everyday internet access becomes available for a lot of people. The business company and banks and other organization which looks for more customers found the internet is the cheapest and the easiest way to increase customers, but they always fear of secret data. Cryptography and steganography appear to solve those fears.

Steganography is Greek word means secret writing. It is the part of information hiding science which focuses on concealing the data (secret data) into an object (covert object). The new object is the stego-object which sends throw unsecure connection to the receiver. Any interception to the stego-object appears normal (without changing in the covert object) to the observer. The receiver can extract the secret data from the stgo-object safety.

Nowadays the digitalized multimedia attracts the attention of information hiding scientist. Steganography can use text, image, audio, video and protocols as covert media. Steganography and cryptography are closely related. Cryptography scrambles messages so they cannot be understood. Steganography on the other hand, will hide the message in a way that hide the first place of the message. This property is so useful especially into battlefields and banks systems. Audio steganography takes advantage of the psycho acoustical masking phenomenon of the human auditory system [HAS]. Psycho acoustical, or auditory, masking is a perceptual property of the HAS in which the presence of a strong tone renders a weaker tone in its temporal or spectral neighborhood imperceptible. This property arises because of the low differential range of the HAS even though the dynamic range covers 80 dB below ambient level [1].
There are many techniques of Audio Steganography can be implemented into audio. They are temporal domain and transformed domain. Several methods of transformed domain contain the frequency domain, wavelet domain, Encoder domain. It is clearly that the tone insertion method comes under transformed domain into frequency domain [2]. It relay into frequency masking property. Any new steganography algorithm should get the three steganography triangle sides imperceptible (security), payload capacity (bit rate) and robustness. The weaknesses of this method are lack of transparency and capacity, but it has good embedded bits rate 250 bps and imperceptible and concealment of embedded data [2].

2. RELATED WORKS

K.Coplan etl published two papers in tone insertion technique [3][4]. They try to improve the capacity of tone insertion method. They use utterances from TIMIT (Texas Instruments Massachusetts Institute of Technology) database as host sample. TIMIT is noise- free database. Firstly they take the utterance "she had your dark suit in greasy wash water all year" which is available as 16 bit sample at the rate 16000 per second. With 208 frames, a random data of 208 bits were embedded. They generated tow tones f0 (set at 1875Hz) and f1 (set at 2625Hz) to embedding bit 0 and bit 1 respectively. They divided host utterances to non overlapped segment. Every segment is 16ms in time. For every frame the compute Fe (frame power) and embedding only one bit into frame. If the embedding bit is 0 the power of the f0 is set to 25% of the fe and the power of f1 is set to 0.001 of f0 and vice versa if the embedding bit is 1 the power of f1 set to 25% of the fe and f0 is set to 0.001 of f1. For recovery the every fame power is computed more over the power p0 and p1 for f0 and f1 respectively is computed. Then calculate the ratio of the power fe/p0 and fe/p1 if the first term greater than the second one then the embedded bit was 0 otherwise the embedded bit was 1.

Secondly they use noise host. It is applied in the Greenflag database consisting of noisy recordings of air traffic controllers, as host or cover audio samples. Successive frames for embedding were overlapped with 50 percent to further increase the payload capacity. After test the imperceptible of hidden data the technique had extended for use in covert battlefield communication in which the hidden information can be another utterance. They use speech utterance " seven one" said by male speaker as covert massage and represented it in GSM half rate (GSM 06.20) coding schema resulting in compact form of 2800 bit. They concatenated two TIMIT utterances as cover audio to accommodate the large covert data. Each utterance with 16 bit samples and 16,000 samples/s. Tones for insertion were selected at frequencies of 687.5 Hz, 1187.5 Hz, 1812.5 Hz, and 2562.5 Hz. These frequencies were either absent or weak in the host frames. One of the frequencies is set to 25% of fe the other frequencies are set to negligible power. To embedding set of (0, 0) set the power of f0 to 25 of the frame power and do the same for (0, 1), (1, 0) and (1, 1) for suitable frequency. For

recovery the receiver know the frequency order, so he can extract the data by computing the minimum power rate to the frame power. Another level of security may add by use frequency hopping in these four frequencies and use 4 bit key for every frame. The second experiment increase payload capacity.

3. PROPOSED TECHNIQUE

In this thesis the levels of the power of the inserted tones to three levels was increased and the carrier tone was decreased to two tones rather than four tones. By using this way two bits can be embedded in the frame. Moreover, we use a third inserted tone as controlled tone (convoy frequency CF) to increase the capacity more than two bits into the frame. Here an English text was used to conceal it into an audio file. So to find the best way to use convoy frequency CF we learnt carefully the ANSII code of the typed characters statistically. we decided to use CF for the 0110 pattern because this pattern appears 26 times in the typing character in ANSII at these letters (a, b, c, d, e, f (twice), g, h, I, j, k, l, m, n, o, v, F, V, X, Y, Z, [, ', &, 6). Moreover we will determine the benefits of CF for that random set in percentage. Note any ANSII character is coded into 8 binary bits, so we fetch the pattern only into the complication of two position(the ANSII position is 0,1,2,3,4,5,6,7) because the S.T table takes pairs of bit any time. Note that the 0110 pattern appears into the vowel letters except u in the lower case and clear that the vowel. letters repeats continually into words. This is raised the feasibility of using the pattern 0110. In the discussion part I study the percentage of using 0110 pattern in random set text written by different people. I use the stego-table (S.T table) that shows the frequencies and the power of the frequencies related to the embedded bits. Figure 1 shows the S.T table. Senders and receivers must know the S.T table in order to conceal and recovery the data safety. (Fe) is the fame energy.

Table 1. The S.T table

Tone frequency	Level of power	Embedded data
F1	1 (25% of fe)	0,1
F2	1 (25% of fe)	1,0
F1 and F2	2 (15% of fe)	1,1
F1 and F2	3 (0,01 of fe)	0,0

In case of (0,1) and (1,0) the power of F1,F2 is raised to 25% of fe respectively(level 1 of power). The F1,F2 together raised to 15% of fe(level 2 of power) when the embedded data is 1,1. The F1,F2 together raised to 0,01% of fe(level 3 of power) when the embedded data is 0,0 is set to level three power. Table 2 shows how to use the CF frequency in order to increase the embedded bits more than two into a frame. Note that any character is eight bits in ANSII and I embedded two bits in a frame so the bits to insert are complication of two.

Table 2. The CF table

Convoy frequency CF	Level of power	Embedded bits
F3	1 (25% of fe)	The following bit is 0110

In the embedded process after convert text data into binary checked, the 0110 pattern into the binary file if it is found raised the CF to 25% of fe. In the recovery process firstly check the CF if it is 25% of fe extract the 0110 pattern else check the f1 power and f2 power according to S.T table.

3.1 The Embedded Algorithm

1. Convert text file into binary file according to ANSII code.
2. Divide the audio file into frame.
3. Computer the power of the fame (fe).
4. Use the S.T table and CF table to raise the suitable frequency related with embedded bits.
5. Go to step3 until the end of the text file.

3.2 The Recovery Algorithm

1. Divide the stego-object into frame.
2. Compute the power of the frame fe.
3. Filter the cf frequency into frame.
4. Check the power of cf
   If it is 25% fe output 0110.
   Go to step 1.
5. Check the power of f1 (f1p)
   If it is 25% of fe
   Output 0,1
   If it is 15% of fe

   Output 1,1
   If it is 0,01% of fe
   Output 0,0
   Else if check f2p
   If it is 25% of fe
   Output 1,0.

6. Go to step 1 for second frame until the end of the audio.

Note: check f1p is enough to determine the hidden data but you can check f2p to be certain.

.

4. EXPERIMENTAL RESULTS and DISCUSSION

To determine the feasibility of using the pattern (0110) . We take random set from different persons(twelve text) the median length of these text is about 840 words(into node bad). These text has been send to specific program in order to determine the size of the text after applied the s.t table and how many times the pattern 0110 appears on that text. Finally we compute the feasibility of using the pattern 0110. Table 3 shows these results.

Table 3. The Feasbility of Using 0110 Pattern

File name	Items of S.T	No-Of-Pattern	Benefit ratio
File 1	7432	502	15.6192%
File 2	16720	964	5.7656%
File 3	10679	1725	16.1532%
File 4	4312	568	13.1725%
File 5	7390	1118	15.1286%
File 6	7495	1101	14.6898
File 7	6772	988	14.5895
File 8	5321	783	14.7153
File 9	5492	720	13.1100
File 10	2115	353	16.6903
File 11	366	26	7.1038
File 12	87	17	19.5402

The best value of the benefit ratio is file10 and file 3. These files include just small letters without number. So to increase the benefit ratio of the pattern detection 0110 the senders must have two different audio. For every audio I change the text file(use two text file length) and the segment length (use three different length).

The first experiment I use 8 sec audio witch available into 16 b/sec sample rate and 16 sample as host audio file. Divide this audio into segments each one is 256 sample (16 m.sec). and use text data with 130 byte to hide it into the audio, Also I use the same file to hide 518 byte in it. (the audio was very small to hide this data). Moreover I do the same experiments but this time I change the segment length to 128 sample (8 m.sec) and 64 (4 m.sec).

The second experiment I use 14 sec audio witch available into 32000 b/sec sample rate and 8 sample as host audio file. Divide this audio into 256 sample length(8 m.sec). Also I change the segment length to 128 sample (4 m.sec) and 64 (4m.sec). Every time I use the same text files above to hide it into this audio. Table 4 shows all experiments.

Table 4. The Feasbility of Using 0110 Pattern

Audio	Text file In byte	Stego-name	Available capacity in byte	Note
Audio 1	130	Steg1_1_256	117	
Audio 1	130	Steg1_1_128	234	
Audio 1	130	Steg1_1_64	468	
Audio 1	518	Steg1_2_256	117	Big text
Audio 1	518	Steg1_2_128	234	Big text
Audio 1	518	Steg1_2_64	468	
Audio 2	130	Steg2_1_256	848	
Audio 2	130	Steg2_1_128	1696	
Audio 2	130	Steg2_1_64	3393	
Audio 2	518	Steg2_2_256	848	
Audio 2	518	Steg2_2_128	1696	
Audio 2	518	Steg2_2_64	3393	

The name of every stego in the stego_name column has three value after the word stego. The first value indicate witch audio has been used. The second value means witch text file is hidden into this stego_file. The last value indicate the segment length.

Note that stego1_2_256 and stego1_2_128 wasn't created because the audio file is too short to hide text file2 (518 byte) but when we use 64 segment length the audio can hide the text file. On the other hand note that stego1_1_256 and stego1_2_64 has been created although the available capacity was too short to hide the certain text file that because the use of the pattern detection.

5. SECURITY METRICS

We use two ways to ensure the security of this method these methods are spectrogram and PSNR,MSE . spectrogram of the audio and stego file it is clearly there are no visual differences between the two figure for all experiments.

Figure 1 to figure 5 shows the spectrograms of the first original audio and it is stego according to experiments

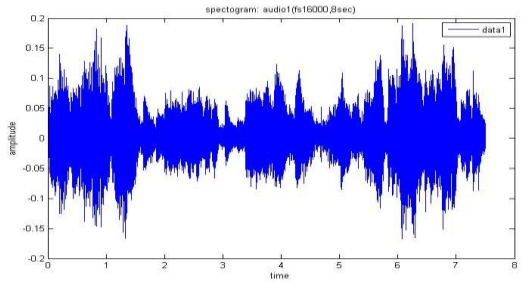


Figure 1 the original audio1

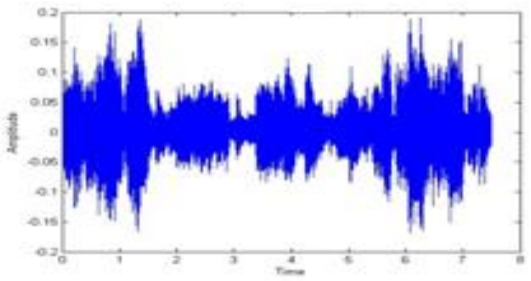


Figure 2 the stego1_1_256

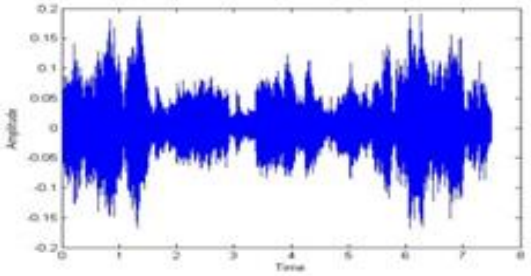


Figure 3 the stego1_1_128

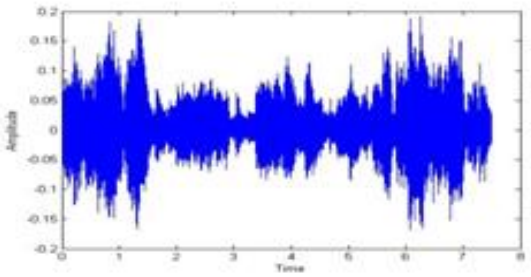


Figure 4 the stego1_1_64

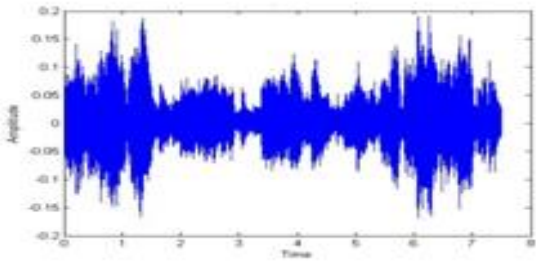


Figure5 the stego1_2_64

Figure 6 to figure 12 shows the spectrograms of the second original audio and it is stego according to experiments.

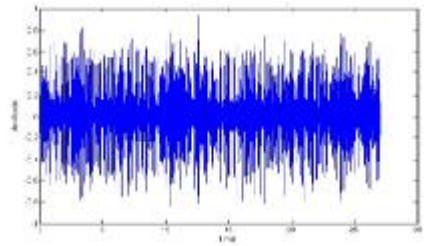


Figure 6 the original audio 2

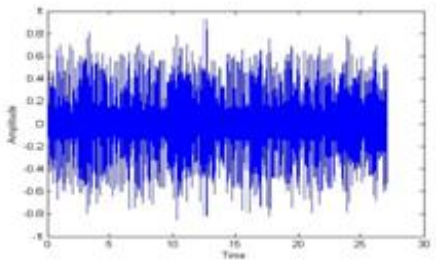


Figure 7 the steg2_1_256

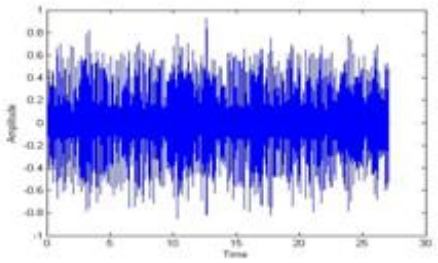


Figure 8 the stego2_1_128

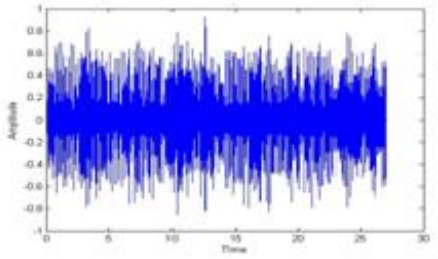


Figure 9 the stego2_1_64

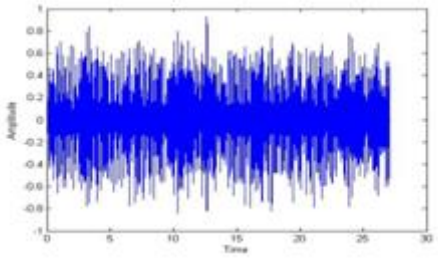


Figure 10 the stego2_2_256

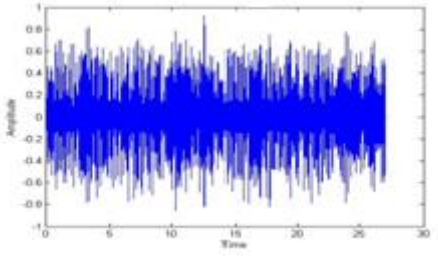


Figure 11 the stego2_2_128

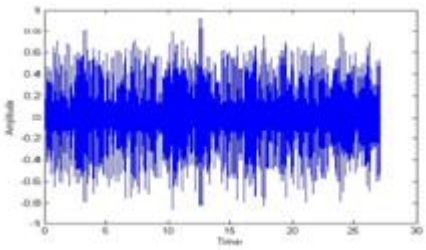


Figure 12 the stego2_2_64

The term peak signal-to-noise ratio (PSNR) is an expression for the ratio between the maximum possible value (power) of a signal and the power of distorting noise that affects the quality of its representation. In statistics, the mean squared error (MSE) of an estimator measures the average of the squares of the "errors", that is, the difference between the estimator and what is estimated. Table 5 shows the PSNR and MSE values for experiments.

Table.5 The PSNR and MSE Values

Audio_name, stego_name	PSNR values	MSE values
Audio1,stego1_1_256	121.6298	0.000003
Audio1,stego1_1_128	125.7655	0.000003
Audio1,stego1_1_64	125.9546	0.000003
Audio1,stego1_2_64	122.0325	0.000003
Audio2,stego2_1_256	103.8790	0.000003
Audio2,stego2_1_128	111.1388	0.000001
Audio2,stego2_1_64	114.8495	0.0000001
Audio2,stego2_2_256	98.2985	0.000010
Audio2,stego2_2_128	102.4410	0.000004
Audio2,stego2_2_64	104.9162	0.000002

From table 5 the PSNR increase when the segment length decrease that because of the signal stationary. When the segment length decrease the audio host is being more stable. So it is clear from table 4 and table 5 that small segment means increasing in security and in capacity. This depend somewhere into the host audio signal .

6. CONCLUSION AND FUTURE WORK

Audio Steganography is very interesting field. The tone insertion technique is imperceptible but the main limitation is the low capacity. Using the proposed method we can increase the capacity to 70% compared with the latest related work.(note that the related work use 2 frequency to insert one bit into frame we use 2 frequency to insert two bit into frame that is 50%. Moreover we use the pattern detection. The feasibility of the patter may arrived to 20% if we applied the mentioned recommendation). The host audio must be carefully choosed. If the host audio has empty sample at the beginning or at the end of it. We have to process it before using it. To guarantee the imperceptible tone and write extraction. We recommended also the host audio must analysis to determine the best segment length. Also use the mid band tone frequency is better. In the future we will concatenate the pattern detection with space character (note that the space character is frequently used). Also we use the human voice as a host audio. Use frequency hopping with known key to deny the frequency tracking and to achieve more security. Increase security by using suitable encryption algorithm for the plain text file before steganography.

7. REFERENCES

[1] E. Zwicker and H. Fastl, 1990 "*Psychoacoustics"*, Spriger-Verlag, Berlin.

[2] Fatiha Djebbar, Beghdad Ayad, Karim Abed Meraim and Habib Hamam, 2012 "Comparative study of digital audio steganography techniques" ,EURASIP Journal on Audio, Speech, and Music Processing.

[3] K Gopalan, S Wenndt, July 8–10, 2004 "Audio Steganography for Covert Data Transmission by Imperceptible Tone Insertion", WOC 2004, Banff, Canada.

[4] K. Gopalan, S. Wenndt, A. Noga, D. Haddad, and S.Adams, , March 2003" Covert Speech Communication Via Cover Speech By Tone Insertion", Proceeding of IEEE Aerospace Conference, Big Sky, Montana.

International Journal of Computer Applications Technology and Research
Volume 6–Issue 6, 259-263, 2017, ISSN:-2319–8656

Text Steganography Using Compression and Random Number Generators

Mohamed Y Elmahi
Elimam Almahdi University
Kosti, Sudan

Talaat M.wahbi
Sudan University of
Science and Technology
Khartoum, Sudan

Mohamed H. Sayed
The National Ribat University
Khartoum, Sudan

Abstract: A lot of techniques are used to protect and hide information from any unauthorized users such as Steganography and Cryptography. Steganography hides a message inside another message without any suspicion, and Cryptography scrambles a message to conceal its contents. This paper uses a new text steganography that is applicable to work with different languages, the approach, based on the Pseudorandom Number Generation (PRNG), embeds the secret message into a generated Random Cover-text. The output (Stego-Text) is compressed to reduce the size. At the receiver side the reverse of these operations must be carried out to get back the original message. Two secret keys (Hiding Key & Extraction Key) for authentication are used at both ends in order to achieve a high level of security. The model has been applied to different message languages and both encrypted and unencrypted messages. The experimental results show the model's capacity and the similarity test values..

Keywords: Text Steganography, Pseudorandom Number Generators (PRNGs), Huffman Compression Algorithm, Cryptography, Capacity ratio, Jaro-Winkler distance

1. INTRODUCTION

Information hiding is a powerful technique used in information security, It takes two general approaches Cryptography and Steganography to hide internet communications [1]. The word steganography comes from two roots in the Greek language, "Stegos" meaning hidden/covered/roof, and "Graphia" simply meaning writing [2]. The history of steganography can be traced back to around 440 B.C.

Steganography is a popular technique of information hiding approaches, the purpose of it to covert communication to hide the existence of a message from a third party. Steganography can be classified into four types image, text, audio and video steganography that is depending on the cover media used to embed secret message [3] (as shown in Figure 1). Due to the significance of the information Cryptography and Steganography are ways of secure data transfer over the Internet [4].

2. RELATED WORKS

Text steganography plays significant role in covert information on Internet. Text steganography although can be broadly classified into three types. Firstly, the Format based, which changes the formatting of the cover-text to hide the data.

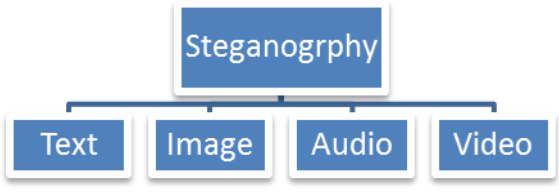


Figure. 1 Steganography Types

Secondly, Random and Statistical generation to avoid comparison with a known plaintext, steganographers often resort to generating their own cover texts. Lastly, Linguistic methods specifically consider the linguistic properties of generated and modified text, in this method a pre-selected synonyms of words are used [3-5].

A lot of studies cover text steganography such as:

Shirali-Shahreza, M.H. and M. Shirali-Shahreza [5] deal with the issue of text steganography, their model focuses on the letters that have points on them (example English Language had two letters i,j. while Arabic language has 15 pointed letters out of its 28 alphabet letters). Point steganography hides information in the points of the letters specifically in the points' location within the pointed letters. After converting the message into bits, if the bit is one the point in the cover text is shifted up, otherwise, the concerned cover-text character point location remains unchanged.

Gutub,A. and M. Fattani. A in [6], "That Benefiting from Shirali-Shahreza [5] proposes a new method to hide information in any letters (Unicode system) instead of pointed ones only. This model uses the pointed letters with extension after the letters to hold secret bit 'one' and the un-pointed Letters with extension to hold secret bit 'zero.

Bhattacharyya, S., I. Banerjee, and G. Sanyal [7] proposes a new method of information hiding in a text by inserting extra blank spaces (one or two spaces) between the words of odd or even size according to the embedding sequence (binary number) of the message.

Banerjee, I., S. Bhattacharyya, and G. Sanyal [8] do same as in [7], except it focuses on the first character of the words in the text cover, if it is a vowel or consonant instead of odd or even size.

Bhattacharyya, S [9]., design a secret key steganographic model combining both text and image first uses a plain text as the cover data and the secret message is embedded in the cover data to form the stego text which in turn is embedded into the cover image to form the stego image. The proposed text steganography scheme has been inspired by the author's previous work [8]. Here data embedding in an image

has been done through Pixel Mapping Method (PMM) within the spatial domain of any gray scale image.

3. HUFFMAN COMPRESSION ALGORITHM

Data compression schemes can be divided into two broad classes: lossless compression schemes, and lossy compression schemes. Lossy compression techniques involve some loss of information. Lossless compression techniques involve no loss of information.

Huffman coding is a lossless data compression. It uses a variable-length code table for encoding a source symbol (such as a character in a file) where the variable-length code table has been derived in a particular way based on the estimated probability of occurrence for each possible value of the source symbol. It was developed by David A. Huffman[10].

The algorithm constructs a tree that is used to represent the characters in the file to be compressed; in the tree (a binary tree) all characters are stored at the tree leaves, each characters has an associated weight equal to the number of times the character occurs in a file. The characters of large weight numbers have less representation bits.

4. RANDOM NUMBER GENERATORS

Random numbers play a significant role in the use of encryption for various network security applications. Random number generators (RNG) are of three types; the first types are the true random number generators (TRNGs) that their output cannot be reproduced. TRNGs are based on physical experiment such as coin flipped 80 times and the Result recorded as binary bit. So it is impossible to generate bit same bit again by using of the same way.

The second types are Pseudorandom Number Generators (PRNG) generates sequences which are computed from an initial seeds, and produces a sequence of output bits using a deterministic algorithm. Typically, PRNG can work by feedback path. PRNG uses the flowing formula:

$$S[i+1] = S[i]*A + b \bmod m; i = 0,1,2,3....$$

$$S[i]; A; B \in \{0,1,2.....m-1\}$$

A; B; m are integer constants.

Third Pseudorandom Number Function (PRF) is used to produce a pseudorandom string of bits of some fixed length such as fixed length keys.

5. THE PROPSED MODEL

The model focuses on text steganography specifically the second type (Random and Statistical generation); the proposed model facilitates text steganography using to be side to side with cryptography to secure sent traffic. The model is divided into two major sites the sender site (Embedding+ Compression) that deals with the embedding processes of the secret message, and the receiver site (Decompression+ Extraction, in reverse order) that deals with the extraction processes to obtain the Secret message safely again. applied to reduce the size. As shown in Figure 2

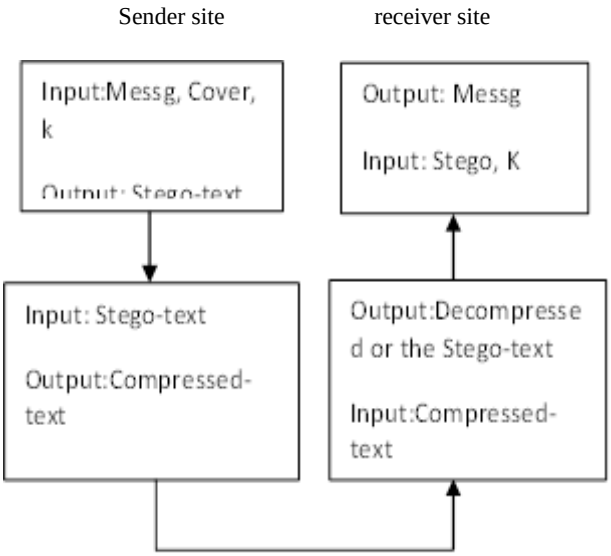


Figure. 2 The Proposed Model

The idea in the Model is to produce random text characters called Cover-text to hide the Secret-Message randomly (different positions) into the Cover-text by the use of the Pseudo Random Generation (PRNG).

The Embedding Algorithm (Sender Site):

-enter the Secret Message.

-calculate the message characters number

-generate the Cover-text from the secret Message and it is more than the message characters number

-enter a number to be the Hiding key.

-from the key generate an array of random number by the equation . A, N Are const.

-generate the binary array (just zeros and ones) by the reminder (mod) of 2 of the above step output numbers.

-generate the binary array until the ones (1s) numbers in the array equal the number of the Secret-Message characters.

-if the binary element equal to 0 writes one character from the Cover-text into the Stego-text, else write from the Secret-Message into the Stego-text.two texts merged into one text file randomly.

-do until the last character in the Secret-Message embedded. The output is a mix of random Cover-text with random position of every single character of the Secret-Message into the new file called Stego-text although it totally seems to be random text.

-Compression process (of Stego-text) it is the last step in the sender site. Figure 2

The Extraction Algorithm (Receiver Site):

- Decompression Stego-text file that is received from sender.

-receive the Decompressed file (or the Stego-text) from Second level of the receiver site.

- enter the extraction key.

-generate array of integers from the key by the same equation in-generate a binary array from the above array of integers by the same way of the first level in the sender site.

-if the binary array element equal 1 write from the stego-text to new text file called The Secret-message, else write one character from the stego-text to new text file called Cover-text2.

-do until the last character in the Stego-text.

Suppose that we have a Message (Hello World) we want to hide by using our model, First extract unique characters from the Message (Helo wrd), Second generate random Cover-text from the unique message characters (oWloWHd W oldHo), Third generate embedding bits from the hiding key (01100101010000100010110011), Fourth use the bits sequence to mix the message and the cover-text to produce Stego-text (oHeWllolWoHd W olWdorHold), fifth compress the Stego-text. The receiver must reverse all these steps to obtain the message again if he has the Extraction key that decided by sender. If the sender decides to send the same message again the cover-text and the stego-text will be different from the above one.

By the unknowing of the message and the cover-text the attackers cannot extract the message from the stego-text, by combining this with encrypted text message

6. RESULTS and DISCUSSION

The different experiments are done to different files of secret message (177, 49 character). The model capacity (which is an ability of a cover media to hide secret information) and similarity (which is the differences between cover text and Stego text) tested. Also Arabic & Encrypted file can be examined (work in this model) too.

The capacity ratio is calculated by dividing the amount of hidden bytes over the size of the cover text in bytes [21].

Capacity ratio = (amount of hidden bytes) / (size of the cover text in bytes)

177/181=0.98 (percentage 98%)

Jaro-Winkler distance for measuring similarity between two strings (s1, s2), it uses as a duplicate detection, Jaro-Winkler value is a ratio between 0 (no similarity) and 1 (an exact match). The Jaro Winkler distance (dj) formula is

$$dj = \frac{1}{3}\left[\frac{m}{|s1|} + \frac{m}{|s2|} + \frac{m-t}{m}\right]$$

Where m is number of matching characters, t is the number of transformation. The match Range computed by

$$matchRange = \left[\frac{max\lfloor|s1|\rfloor, \lfloor|s2|\rfloor}{2}\right] - 1$$

Table 1: Result Table

Mess	Cover	Stego	Comps	Ca-Ratio	JW-Ratio
49	47	96	444	1.04	0.5843
49	40	89	407	1.23	0.5622
49	35	84	381	1.40	0.5714
177	181	358	1599	0.98	0.6397
177	169	346	1557	1.05	0.6260
177	165	342	1547	1.07	0.6144

Table 2: Average Result Table

Mess	Cover	Stego	Comps	Ca-Ratio	JW-Ratio
49	41	90	411	1.22	0.57
177	172	349	1568	1.03	0.63

In all the tables above from left to right columns Mess-No are the values of the message characters number Cov-No are the values of the cover characters number, by adding these two values that gave us the Stego-No values, after the compression of the Stego-No files the values in bits found in the column Comp-Bits, Ca-Ratio are the capacity test values. JW-Ratio are the similarity test values between cover text files and stego text files.

If the values in the column Mess-No multiplied by 8 (suppose it is ASCII), that gave us the message size in bits instead of characters number (e.g. 49x8=392 or 177x8=1,416), if the comparison made between the those values (message value in bits) and the values of the column Comp-Bits in the same row, you will find small differences may be less or greater than the message bits value. Those differences are due to two reasons, first the random generation of the cover characters, second the Compression Algorithm that depends on the character frequency in the file, repeated characters which represented as less bits.

7. CONCLUSION AND RECOMENDATION

The purpose of this study is to conceal the sensitive information from an unauthorized use by hiding the Secret Message into Cover-Text generated randomly with the ability of extracting the Secret Message again. The study utilizes a compression algorithm as the next step that adds a good feature to the model. The essential role of a good compression algorithm is to reduce the size of the files, the increasing demand for the compressed data is to hasten the transfer rate

and operation, after the compression process is done the output of the compressed file is different from the original file itself, but it is near to the original Secret message's size.

In other site Text Steganography (in the model) is measured, such as the number of the Cover-Text characters that is used to randomize the message inside it, it used as rate of security (the larger number of the Cover-Text the better security indicator). Although the Embedding operation is faster(by embedding 8bit at a time instead of 2 or 3bit at a time than the other).

if the sender decided to send the same message more than one the output of the stego-text will be differ (due to of the random cover-text generating, and although Huffman Compression algorithm).

The model eliminates the overhead of find suitable cover-text to hide different messages (difference in type or size), the sender and the receiver exchanges stego-text files only any third party cannot obtain the secret messages again because of hidden cover-text files that produced and without having the extraction keys. The model does not use cover text data set.

8. FUTURE WORKS

Message permutation before the sending can be applied to the model, also we can embed the last character instead of the first one

First alter the design to accept different type of messages such as images, audio, etc. Second apply the random embedding operations to the binary representation (e.g ASCII) outputs of the secret message and the cover-text to hide the data more deeply. The model can also be implemented to the data transferred through network or internet whether it is plaintext or encrypted data.

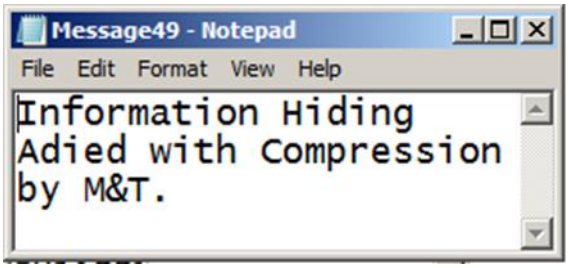


Figure. 3 49 Characters Message

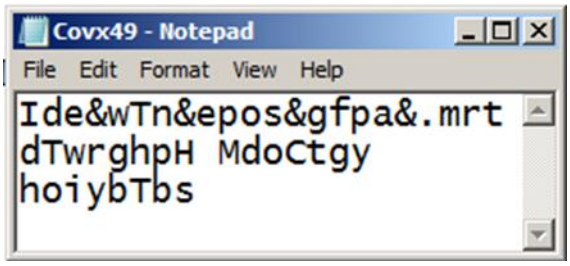


Figure. 4 Cover-text of the Message Above.

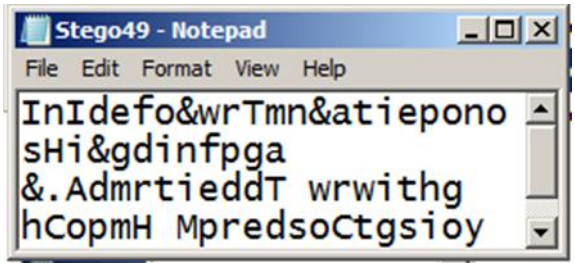


Figure. 5 Stego text

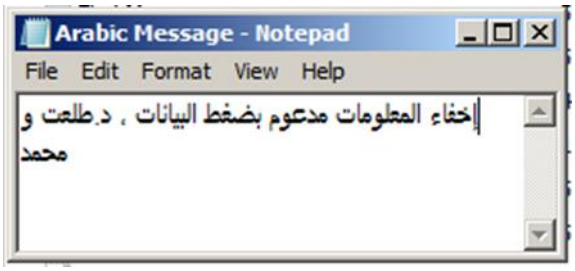


Figure. 6 Arabic messages

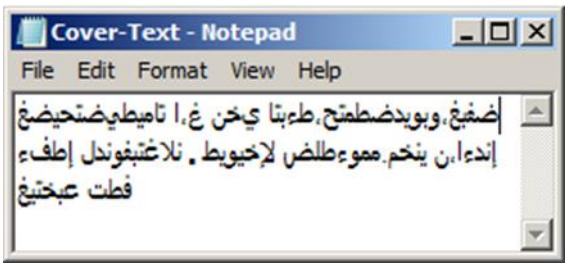


Figure. 7 Arabic Cover-text

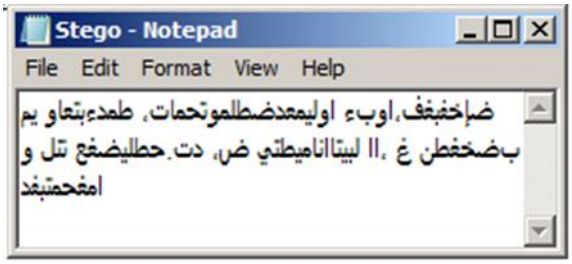


Figure. 8 Arabic Stego- text

9. REFERENCES

[1] Al-Najjar, A.J. 2008, *The decoy: multi-level digital multimedia steganography model*. in *WSEAS International Conference. Proceedings. Mathematics and Computers in Science and Engineering*. World Scientific and Engineering Academy and Society.

[2] Krenn, R., 2004, *Steganography: Implementation & Detection.* found online at<http://www.krenn.nl/univ/cry/steg/presentation/2004-01-21-presentation-steganography.pdf,

[3] Isbell, R., 2002, *Steganography: hidden menace or hidden saviour*. Steganography White Paper.

[4] Agarwal, M., 20135(1) TEXT STEGANOGRAPHIC APPROACHES: A

COMPARISON. International Journal of Network Security & Its Applications,.

[5] Shirali-Shahreza, M.H. and M. Shirali-Shahreza, 2006, *A new approach to Persian/Arabic text steganography*. in *Computer and Information Science, 1st IEEE/ACIS International Workshop on Component-Based Software Engineering, Software Architecture and Reuse. ICIS-COMSAR* .

[6] Gutub, A. and M. Fattani, 2007, *A novel Arabic text steganography method using letter points and extensions*. in *WASET International Conference on Computer, Information and Systems Science and Engineering (ICCISSE), Vienna, Austria*.

[7] Bhattacharyya, S., I. Banerjee, and G. Sanyal, 2010 *A novel approach of secure text based steganography model using word mapping method (WMM).* International Journal of Computer and Information Engineering, **4**(2): p. 96-103.

[8] Banerjee, I., S. Bhattacharyya, and G. Sanyal, 2011 *Novel text steganography through special code generation*. in *Proceedings of International Conference on Systemics, Cybernetics and Informatics (ICSCI-2011), Hyderabad, India*.

[9] Bhattacharyya, S.,2011, *Data hiding through multi level steganography and SSCE.* Journal of Global Research in Computer Science.

[10] *Huffman Compression Algorithm*. [cited 2014 2/3/].

[11] *A. Haidari, A. Gutub, K. A. Kahsah, and J. Hamodi, July 2012 "Improving security and capacity for Arabic text steganography using "*kashida" extensions," *2009 IEEE/ACS Int. Conf. on Computer Systems and Applications, 2009, pp. 396-399.*Sharon Rose Govada, Bonu Satish Kumar, Manjula Devarakondaand Meka James Stephen:Text Steganography with Multi level Shielding. IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 3,

International Journal of Computer Applications Technology and Research
Volume 6–Issue 6, 264-268, 2017, ISSN:-2319–8656

An e-Readiness Assessment of ICT Integration in Public Primary Schools in Kenya Case of Nyeri County.

Rachael Njeri Ndung'u
Department of Computing Sciences
University of Embu, Kenya

Florence Maweu,
Department of Computing and Information Technology,
Kirinyaga University, Kenya

Joyce M. Mwenja
Department of ICT
Kirinyaga University, Kenya

Abstract: In primary schools, information and communication technologies (ICT) are widely seen as tools for enhancing learning. This expectation increases their rapid diffusion and adoption throughout developing countries. Despite the strong emphasis given to ICTs in education, little has been done to evaluate e-readiness as a factor that influences ICT integration in primary schools in Kenya. This study sought to evaluate e- readiness of ICT integration in public primary schools in Nyeri Central Sub-County, Kenya. The study adopted descriptive survey design. The targeted population was 376 subjects comprising of 375 teachers in the 23 public primary schools in Nyeri Central Sub-county and one Sub-county director of education. 11 schools were sampled through simple random sampling technique. The SCDE and the head-teachers of every school sampled were purposively selected. 66 teachers were randomly sampled, making the sample size of 78 respondents. Data collection was done using questionnaires, interview and observations. There was a response rate of 100%. The results indicated moderate readiness in adopting ICT in public primary school in Nyeri Central Sub-county. This would be enhanced if the government would expedite the process of procurement and installation of the digital hardware and software in the schools and also, plan on how all public primary teachers would be re-trained on ICT application on the primary education curricula.

Keywords: ICT; e-Readiness; integration; primary schools; education; e-readiness

1. INTRODUCTION

The reasons for introducing Information and Communication Technology (ICT) in schools are four-fold: technology innovations, globalization of economy and information, knowledge based economy and society and increasing demand for education (Temba, 2013). The role of ICT in promoting economic growth and development has gained prominence globally. Economies are been transformed from industrial to knowledge-based where knowledge is recognized as a driver of productivity and economic growth (OECD, 2004; World Bank, 2008). It's expected that adopting and using ICT in schools would lead to significant expansion of education and academic outcomes which are beneficial to both teachers and students. When used appropriately, ICT can help to support the value of education by increasing access to education opportunities, raising quality of education by making learning and teaching an active process connected to real life (Zaman, Shamim & Clement, 2011).

ICT integration to curricula requires physical infrastructure, technology expertise and emotional readiness for those who are bound to implement it. As a first, the Kenyan government is making efforts to initiate e-learning in public primary schools by connecting all public primary schools to the national electricity grid. These efforts are aimed to prepare the schools to adopt ICT.

Despite the strong emphasis given to ICTs in education by even publishing Kenya national ICT policy (2006), little has been done to evaluate e-readiness as a factor that would influence ICT integration in primary schools in Kenya. Minimal e-readiness evaluation or lack of it would mean starting the ICT integration process without facts of whether the schools have capacity for integration or whether the process would successfully complete as was intended. It was on this background that this study sought to investigate the question: what was the level of physical and technological readiness of ICT integration in public primary schools in Nyeri Central Sub-county?

2. OBJECTIVES OF THE STUDY

The purpose of the study was to investigate and evaluate the various factors and extent to which public primary schools in Nyeri Central Sub-county in Kenya were prepared for ICT integration to the education system (curricula and administrative activities).

The general objective of this study was to evaluate the e-readiness of ICT integration in public primary schools in Nyeri Central Sub-county. In order to get more on the e-readiness, the general objective was further divided into more specific objectives, which were:-

To evaluate the extent to which public primary schools' institutional infrastructure can accommodate ICT facilities;

To establish the level of preparedness of teachers in using ICT in teaching and learning in the classroom;

To evaluate other factors that would hamper effective ICT integration in public primary schools in Nyeri Central Sub-county.

The research was guided by the following questions:

To what extent are public primary schools ready to integrate ICT in their institutions?

What is the level of preparedness of teachers in using ICT on teaching and learning methods?

What are other factors that would hamper effective ICT integration in public primary schools in Nyeri Central Sub-county?

3. LITERATURE REVIEW

3.1 Technology Readiness

Ouma, Awuor, & Kyambo (2013) did a study on e-learning readiness in public secondary schools in Kenya, and he sought to check e-readiness in terms of teachers' competencies, perception and attitude towards ICT. The study was based on schools that the government had selected to fund and support ICT infrastructure development and teacher's training in ICT. Ouma et al. (2013) in their study concluded that teachers were moderately ready for e-learning and that there were individuals who may need to be acculturated into the e-learning system before they were said to be at the expected state of readiness for e-learning. They also found out that policy makers and other education stakeholders had a crucial role to play in enhancing greater engagement in a technology-driven teaching-learning environment. However, while Ouma et al. (2013) study was based on five secondary schools that were government sponsored to improve e-learning, this current study is evaluating e-readiness in public primary schools which are not receiving special ICT funds to improve e-learning from the government.

3.2 Institutional Infrastructure Availability and Readiness

In a study done in Kenya to evaluate the level of ICT investment in the education sector, Hennessey et al. (2010), while reviewing literature stated that the level of investment in ICT in education in Kenya reflects the recognition in the national ICT policy of the need for Public-Private Partnership (PPP) in addressing key development challenges in the country. Towards strengthening adoption and use of ICT in the education sector in Kenya, an ICT unit has been established at the Ministry of Education, Science and Technology (MoEST)'s head office to ensure systematic efforts are made. In this respect, the Government of Kenya, through MoSET plays a coordinating, overseer and mobilization role in bringing together key stakeholders in the ICT in education sector.

3.3 Institutional Infrastructure Availability and Readiness

In a study done in Kenya to evaluate the level of ICT investment in the education sector, Hennessey et al. (2010), while reviewing literature stated that the level of investment in ICT in education in Kenya reflects the recognition in the national ICT policy of the need for Public-Private Partnership (PPP) in addressing key development challenges in the country. Towards strengthening adoption and use of ICT in the education sector in Kenya, an ICT unit has been established at the Ministry of Education, Science and Technology (MoEST)'s head office to ensure systematic efforts are made. In this respect, the Government of Kenya, through MoSET plays a coordinating, overseer and mobilization role in bringing together key stakeholders in the ICT in education sector.

3.4 Perceived Gap

Reviewed studies have shown the need for adopting technology in the education sector. This adoption has to take place in all levels of education, starting from primary schools to higher institutions of learning. The ICT integration process are under-way but there is still unanswered question, are primary schools ready to adapt to technological aspects of teaching and learning? Could there be issues that are supposed to be addressed before integration starts? A study conducted by Hennessey et al. (2010) pointed out a summary of factors derived from a survey previously done. They noted that majority of teachers were ill equipped to effectively integrate ICT in classroom. The main challenges for teachers interviewed were lack of adequate number of computers, educational applications, training, policy and strategy on how integration should be done.

Not much is documented about ICT integration readiness in primary schools in Kenya. It is in this context the researcher wished to investigate the extent to which teachers were prepared for ICT in the classroom, the ICT infrastructure availability and access and generally how well prepared the public primary schools in Nyeri Central Sub-county are in adopting ICT in their education system.

4. RESEARCH METHODOLOGY

4.1 Research Design

The research was guided and conducted using the descriptive survey research design. Descriptive survey research designs are used in preliminary and exploratory studies to allow researchers to gather information, summarize, present and interpret for the purpose of clarification (Orodho, 2002).

4.2 Geographical Description of the Study

This research study was conducted in Nyeri Central Sub-County, Kenya. Nyeri central sub-county is located in Central region, in Nyeri County, Kenya. Nyeri County constitutes six sub-counties; Nyeri Central, Othaya, Tetu, Kieni, Mathira and Mukurwe-ini. The Nyeri Central Sub-county has 23 public primary schools. A sub-county director of education is in charge of education in the sub-county. All primary school heads in the sub-county account to the SCDE.

4.3 Target Population

The study targeted 375 teachers and one (SCDE) in Nyeri Central Sub-county. This gave a target population of 376 subjects. The SCDE gave the progress of e-learning readiness in the sub-county; the teachers who participated gave the information on their e-readiness, how they utilize computer technology and any challenges faced in their individual schools.

4.4 Sample Size and Sampling Technique

This study used simple random sampling method to sample 11 schools, which were used as representative of the population of public primary schools in Nyeri Central Sub-county. In every school that was sampled the head-teacher was purposively selected and six (6) other teachers were randomly sampled. The Sub-county director was also purposively sampled making the sample size of [1+11+ (6*11)] =78 respondents. The seven teachers sampled per school were deemed a well representative sample.

4.5 Data Collection Procedures

The research study was a quantitative research. Quantitative data collection techniques used includes: online surveys,and questionnaires.

Methods of Data Analysis

The primary data, after collection, was processed and analyzed. This was essential for a scientific study and for ensuring that we have all relevant data for making contemplated comparisons and analysis. This involved processing of data, which consisted of editing, coding,

classification and tabulation of collected data so that it was amenable to analysis. Percentages, frequency distributions were used to analyze the collected data with an aid of Statistical Package of Social Sciences (SPSS v20.0).

5. DATA PRESENTATION & ANALYSIS

5.1 Response Rate

There were 77 questionnaires that were handed to all respondents. The researcher visited all eleven schools sampled and sought audience first to the head of schools. This was important to the researcher to explain the reason of the research and that the respondents been sought were teachers in those schools, the head teacher inclusive. Then the head of schools distributed the questionnaires randomly to the teachers who duly filled the forms, as the researcher waited. This method of questionnaire distribution gave a 100% response rate. All the 77 questionnaires were filled up. There was also a scheduled interview with the sub-county director of education, which took place at his office. The interview was scheduled for 15 minutes. The response was also very positive since he answered all questions asked and even gave more explanations as of ICT status in the sub-county.

6. SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

The study was to evaluate e-readiness of ICT integration in public primary schools in Nyeri Central Sub-county. Integrating ICT in primary education sounds an appealing method to address access and quality education challenges facing Kenyan education system. This study therefore was investigating e-readiness in very specific areas: institutional infrastructure development, teachers' competencies, funding and effects of governing ICT policies.

On the area of institutional infrastructure, it was found that despite all sampled schools having permanent buildings and electricity, the other key issues like security proofing within the schools premises has to be addressed for a successful ICT adoption. The issues that the research brought out as gaps were lack of ICT facilities (computers, internet, and syllabus content), lack of security, lack of buildings to secure computers and laptops and lack of ICT support staff.

It was found out that the teachers in Nyeri Central Sub-county were in the process of preparing themselves for ICT integration. Majority of the respondent showed regular computer access and their ICT skills were also averagely fair with majority of respondents aged below 40 years having attended the basic ICT classes. It was found-out that none of the respondents had ICT pedagogical integration training.

On other e-readiness considerations, the government would be funding the ICT project for all public primary schools in the country, but it was noted from the research done that other stakeholders like the parents of those schools should fund support and maintenance of the ICT facilities. It was also noted that the ICT support staff, should be government employees and not school board of management employees.

Another consideration was on policy, where it was noted that the national ICT policy (2006) has not taken affect as the guiding principal in schools, where the schools were supposed to demonstrate their ICT preparedness regarding infrastructure development, human resource development. It was found out this was because of lack of funding from the government.

Success in integrating ICT in a classroom setting is heavily dependent on teachers, investments in infrastructure to support a curriculum, the formation of practical ICT skills as well as pedagogical initiatives such as computer-assisted instruction. Evaluating e-readiness of ICT integration in public primary schools was guided by national ICT policy (2006), which acknowledges specific challenges as lack of policy and regulatory framework, inadequate infrastructure and insufficient skilled human resources.

6.1 Discussions

6.1.1 Institutional infrastructure

The findings from the study demonstrated that institutional infrastructure development requires urgent improvement in order to adequately support ICT facilities like cabling, that would in turn enhance connectivity of the network and internet, providing secure environment for e-learning. It was notable that all schools sampled demonstrated availability of electricity and permanent buildings. It was also demonstrated that, 62.3% of the respondents felt parents should fund security of the ICT infrastructure as stakeholders of the schools while the government should bear the cost of constructing computer rooms.

6.1.2 Teachers competency

The technical experience and computer literacy among teachers was found to be fair since 46.8% (36) of respondents access computers regularly. None of the sampled teachers demonstrated training on ICT application in curricula. The study however, revealed that teachers' attitude towards ICT was very positive and they were willing to acquire computing skills if at all the government would expedite the ICT pedagogical integration training sessions.

6.1.3 Other factors hampering e-readiness

Other factors that were brought out by the study that would enhance ICT integration to primary education system were funding and implementation policies. Other stakeholders like the parents have a responsibility of making sure that ICT integration to primary education is successful. Their participation would be in making sure that the installed ICT infrastructure and accessories were secure and well maintained as was demonstrated by 62.3% (48) and 68.8% (53) respondents respectively. This participation means they are expected to fund maintenance of the ICT facilities, while the government funds the entire installation of ICT facilities and mapping them to curriculum taught so as to take advantage of the technology.

The study also revealed that 46.8% (36) respondents were not aware of the Kenya national ICT policy (2006) and how it affects ICT integration in the primary education. There was mixed reactions from respondents in relation to the government policy of introducing laptops to junior primary classes. As was demonstrated by 56% (44) of respondents felt it was a good policy and 18.18% (14) respondents felt it a very poor policy citing lack of prioritizing teachers training on the side of the government. The researcher observed failure of the government through MoSET to sensitize the schools on the various polices they had formulated and also they had failed to follow the progress of certain issues raised by the policies at the school level.

Considering the responses of the respondents who felt laptops for junior primary pupils was misplaced, suggestions like computer training in primary schools should start at higher classes like standard six, seven and eight and not standard one, two and three, were fronted, giving reasons that junior pupils were yet to value laptops as serious gadgets

for learning with, and often they would take them as playing toys.

6.2 Conclusions

In order to know whether the public primary schools in Nyeri Central Sub-county are ready for ICT-based teaching and learning, there was a need for an evaluation of e-readiness. From the findings of the study it can be concluded that public primary schools in Nyeri Central Sub-county are reasonably ready to integrate ICT with the primary school curricula despite many challenges.

6.2.1 Institutional infrastructure

There is an urgent need to improve institutional infrastructure development so as to support ICT facilities like cabling that would enhance connectivity of the network and internet and provide a secure environment for e-learning. Individual schools should take initiative to make such developments. Types of buildings, availability of electricity and security within schools were prerequisite measures of e-readiness as far as infrastructure development is concerned, and as the national ICT policy (2006) anticipates.

6.2.2 Teachers competency

The findings of the study revealed that primary school teachers were fairly versed with computer literacy, with those aged below 39 years having undergraduate degree as their highest qualification. This illustrated that teachers are e-ready even though they scored nil on ICT application into curricula. It was also revealed that teacher's attitude towards ICT was very positive. However, their computer skills could not be translated to teaching with and through technology. There is a dire need for these teachers to be retrained on ICT application in the classroom. MoSET should plan for the pedagogical training sessions of all primary school teachers as a way of improving the numbers of skilled human resource whose mandate would be to implement teaching and learning through ICT in classrooms.

6.2.3 Other factors hampering e-readiness

As the government promised the acquisition of both hardware and installation of digital content, it should show a deliberate process to expedite the procurement process. Further, the study has shown that policy makers and other education stakeholders have a crucial role to play in enhancing greater engagement in a technology-driven teaching-learning environment. As Hennessey, et al. (2010) observed in their study, the level of investment in ICT in education in Kenya reflects the recognition in the national ICT policy of the need for Public-Private Partnership (PPP) in addressing key development challenges in the country.

Finally, to successfully integrate ICT in public primary schools the government should hasten to establish an ICT integration framework to guide the integration process and further, review primary education curriculum to include technological aspects that would assist efficient and effective use of ICT in teaching and learning in the classroom. As an urgent government policy, to build capacity needed for a smooth implementation process, teachers should retrained on how to apply ICT in teaching.

6.3 Recommendations and further work

It can be recommended that all public primary schools infrastructural facilities need to be improved to accommodate installation of ICT facilities. This should be a joint venture of the schools BOM and other schools' stakeholders.

Teachers service commission (TSC), should plan to retrain all public primary school teachers on the use of ICT in the classrooms so as to support teachers implement the same swiftly.

There is need to review the entire primary education curricula so as to take advantage of new technological aspects for all classes. The digital content as the software to be installed on the computers should bear the revised curricula. KICD are responsible for the revision of curricula and are the same who are preparing the digital content to be used. Until when this content will be installed in schools would on know it merits. It would be too early to criticize the digital content.

The government through MoSET should hasten the procurement of ICT devices their installations and also installation of digital content (software) for all public primary schools.

Since the study covered only public primary schools in Nyeri Central Sub-county, and e-learning is been introduced in the whole country of Kenya, further study of the level of e-readiness should be extended to all primary schools in the country both public and private.

7. REFERENCES

[1] Andoh, B. 2012. An Exploration Of Teachers' Skills, Perceptions And Practices Of ICT In Teaching And Learning In The Ghanaian Second-Cycle Schools, Contemporary Educational Technology 3(1), 36-49

[2] Balanskat, A., Blamire, R., & Kefala, S. 2006. The ICT impact report. European Schoolnet.

[3] Cassim, K. M., & Obono, S. E. 2011. On the factors affecting the adoption of ICT for the teaching of word problems. In Proceedings of the World Congress on Engineering and Computer Science (Vol. 1, pp. 19-21).

[4] Chigona, A., & Chigona, W. (2010, June). An Investigation of Factors affecting the Use of ICT for Teaching in the Western Cape Schools. In ECIS.

[5] Chrysostomou, M. and Mousoulides, N. 2009. Teachers' Beliefs about the Adoption of New Technologies in the Mathematics Curriculum. Proceedings of CERME 6, pp. 1270-1279

[6] Ford, D. M. 2007. Technologizing Africa: On the bumpy information highway. Computers and Composition, 24(3), 302-316.

[7] Hennessy,S. Harrison,D. and Wamakote,l. 2010. Teacher actors Influencing Classroom Use of ICT in Sub-Saharan Africa: Itupale Online Journal of African Studies, 2 (2010) 39-54 39.

[8] Hennessy, S., Ang'ondi, E., Onguko, B., Namalefe, S., Harrison, D., Naseem, A., & Wamakote, L. 2010. Developing the Use of Information and Communication Technology to Enhance Teaching and Learning in East African Schools: Review of the Literature: The University of Cambridge. Aga Khan University, Nairobi Kenya.

[9] Jones, A. 2004. A Review of the Research Literature on Barriers to the Uptake of ICT by Teachers. UK: Becta.

[10] Kashorda, M., Waema, T., Omosa, M., & Kyalo, V. 2007. E-Readiness Survey of Higher Education

Institutions in Kenya: A Study Funded by Partnership for Higher Education in Africa.

[11] Khan, H. Hasan, M. & Clement, K. 2012. Barriers to the introduction of ICT into education in developing countries: the example of Bangladesh International Journal of Instruction, 5 (2) 61-80

[12] Kothari, C.R. 2004. "Research Methodology Methods & Techniques", Second Edition, New Delhi: New Age International publisher, P.55.

[13] Kumar, N., Rose, R.C. and D'SIiva, J.L. 2008. Teachers' Readiness to Use Technology in the Classroom: An Empirical Study. European Journal of Scientific Research. 21(4), pp. 603-616.

[14] Lumumba, P. 2007. A Survey of Challenges Facing E-learning at Public Secondary Schools: A Case Study of the NEPAD Pilot Project in Schools in Kenya. Unpublished M.ED Project of the Catholic University of East Africa.

[15] Maruti, S. J. 2011. E-learning readiness among public primary teacher training colleges in Kenya (Doctoral dissertation).

[16] Mulwa, A. S., & Kyalo, D. N. 2013. The Influence of Principals', Teachers' and Students' attitude On Readiness To Adopt E-learning In Secondary Schools in Kitui District, Kenya. European Scientific Journal, 9(5).

[17] Ncunge, D. Sakwa, M & Mwangi, W. 2012. User's perception on ICT adoption for education support in schools: A survey of secondary school teacher's in Thika District Kenya, International Journal of Humanities and Social Science 2(10) 17-29

[18] Nzuki, D. 2014. Factors Influencing ICT Integration in Teaching-A literature Review.

[19] OECD & World Bank. 2004. Korea and the Knowledge-Based Economy: Making the Transition.

[20] Orodho J. A. 2002. Techniques of Writing Research Proposals and Reports in education and Social Sciences. Nairobi: Masola Publishers.

[21] Ouma, G. O., Awuor, F. M., & Kyambo, B. 2013. Evaluation of E-Learning Readiness in Secondary Schools in Kenya.

[22] Pelgrum, W. J., & Law, N. 2003. ICT in education around the world: Trends, problems and prospects. Unesco, International Institute for Educational Planning.

[23] Ncunge, D. Sakwa, M & Mwangi, W (2012). User's perception on ICT adoption for education support in schools: A survey of secondary school teacher's in Thika District Kenya, International Journal of Humanities and Social Science 2(10) 17-29

[24] Nzuki, D. (2014). Factors Influencing ICT Integration in Teaching-A literature Review.

[25] OECD & World Bank. (2004). Korea and the Knowledge-Based Economy: Making the Transition.

[26] Orodho J. A. (2002). Techniques of Writing Research Proposals and Reports in education and Social Sciences. Nairobi: Masola Publishers.

[27] Ouma, G. O., Awuor, F. M., & Kyambo, B. 2013. Evaluation of E-Learning Readiness in Secondary Schools in Kenya.

[28] Pelgrum, W. J., & Law, N. 2003. ICT in education around the world: Trends, problems and prospects. Unesco, International Institute for Educational Planning.

[29] Sang, G., Valcke, M., Van Braak, J. and Tondeur, J. 2009. Factors support or prevent teachers from integrating ICT into classroom teaching: A Chinese perspective. Proceedings of the 17th International Conference on Computers in Education. Hong Kong: Asia-Pacific Society for Computers in Education. pp. 808-815.

[30] Temba, John, ed. 2013. 'ICT Policy and Framework, Four volumes, Ministry of Education, Science and Technology': ICT Vision.

[31] Tondeur, J., Valcke, M., & Van Braak, J. 2008. A multidimensional approach to determinants of computer use in primary education: Teacher and school characteristics. Journal of Computer Assisted Learning, 24(6), 494-506.

[32] Zaman, M. Shahim R., & Clement K. 2011. Trends and issues to integrate ICT in teaching learning for the future world of education. International journal of Engineering & Technology 11(3) 114- 119.

[33] Zhang, J. 2004. Using ICT to prepare learners for the 21st Century: The perspectives of the eastern APEC economies. APEC Summit on Educational Innovation: "Striking Balance: Sharing Practice from East and West" Beijing.

International Journal of Computer Applications Technology and Research
Volume 6–Issue 6, 269-272, 2017, ISSN:-2319–8656

Web Mining and Qualities of a Website Design to Be Evaluated for Customer Browsing Behavior: A Review

Sunil B. Joshi
Assistant Professor,
SIMCA,Narhe, Pune,
MH, 411041

Dr. Shivaji D. Mundhe
Professor & Director (MCA),
SIMCA, Narhe, Pune,
MH, 411041

Abstract: Mining the web is defined as discovering knowledge from hypertext and World Wide Web. The World Wide Web is one of the longest rising areas of intelligence gathering. Now a day there are billions of web pages, HTML archive accessible via the internet, and the number is still increasing. However, considering the inspiring diversity of the web, retrieving of interestingness web based content has become a very complex task. In this paper researcher has done a review of web mining concepts with techniques and the some of the qualities of a website design to be evaluated for customer browsing behavior.

Keywords: Web Mining, Web Content Mining, Web Usage Mining, Web log Mining, Browsing Behavior, website design qualities

1. INTRODUCTION

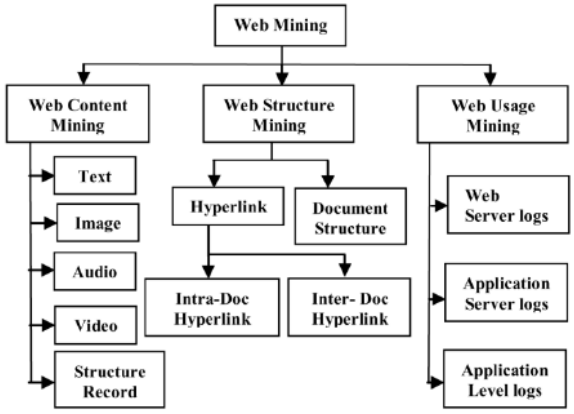
Web mining is the application of data mining techniques to extract knowledge from web data, including web documents, hyperlinks between documents, usage logs of web sites, etc. Data mining is the process which automates the extraction of predictive information, discovers the interesting knowledge from large amounts of data stored in databases, data warehouses or other information repositories. The WWW continues to grow at a wonderful rate as an information gateway and as a medium for conducting business. Web knowledge mining has been widely used in the past for analyzing huge collections of data, and is currently being applied to a variety of domains. Based on several research studies web mining can be broadly classified into three domains: content, structure and usage mining. Web content mining is the process of extracting knowledge from the content of the actual web documents (text, content, multimedia, etc.). Web structure mining is targeting knowledge from the Web structure, hyperlink references and so on. Web usage mining attempts to discover useful knowledge from the secondary data obtained from the interactions of the users with the Web.

2. WEB MINING TAXONOMY

Web mining can be broadly divided into three distinct categories, according to the kinds of data to be mined.

2.1 Web Content Mining

Web content mining is the process of extracting useful information from the contents of web documents. Content data is the collection of facts a web page is designed to contain. It may consist of text, images, audio, video, or structured records such as lists and tables. Application of text mining to web content has been the most widely researched. Issues addressed in text mining include topic discovery and tracking, extracting association patterns, clustering of web documents and classification of web pages. Research activities on this topic have drawn heavily on techniques developed in other disciplines such as Information Retrieval (IR) and Natural Language Processing (NLP). While there exists a significant body of work in extracting knowledge from images in the fields of image processing and computer vision, the application of these techniques to web content mining has been limited.



Web Structure Mining

The structure of a typical web graph consists of web pages as nodes, and hyperlinks as edges connecting related pages. Web structure mining is the process of discovering structure information from the web. This can be further divided into two kinds based on the kind of structure information used.

Hyperlinks

A hyperlink is a structural unit that connects a location in a web page to a different location, either within the same web page or on a different web page. A hyperlink that connects to a different part of the same page is called an intra-document hyperlink, and a hyperlink that connects two different pages is called an inter-document hyperlink.

Document Structure

In addition, the content within a Web page can also be organized in a tree structured format, based on the various HTML and XML tags within the page. Mining efforts here have focused on automatically extracting document object model (DOM) structures out of documents.

2.2 Web Usage Mining

Web usage mining is the application of data mining techniques to discover interesting usage patterns from web usage data, in order to understand and better serve the needs of web-based applications. Usage data captures the identity or origin of web

users along with their browsing behavior at a web site. Web usage mining itself can be classified further depending on the kind of usage data considered:

Web Server Data

User logs are collected by the web server and typically include IP address, page reference and access time.

Application Server Data

Commercial application servers such as Weblogic, StoryServer, etc have significant features to enable E-commerce applications to be built on top of them with little effort. A key feature is the ability to track various kinds of business events and log them in application server logs.

Application Level Data

New kinds of events can be defined in an application, and logging can be turned on for them - generating histories of these events. It must be noted, however, that many end applications require a combination of one or more of the techniques applied in the above the categories.

3. KEY CONCEPTS OF WEB MINING

3.1 Ranking Metrics—for Page Quality and Relevance Searching the web involves two main steps

Extracting the pages relevant to a query and ranking them according to their quality. Ranking is important as it helps the user look for "quality" pages that are relevant to the query. Different metrics have been proposed to rank web pages according to their quality.

We briefly discuss two of the prominent ones.

PageRank

PageRank is a metric for ranking hypertext documents based on their quality. Page, Brin, Motwani, and Winograd (1998) developed this metric for the popular search engine Google4 (Brin and Page 1998). The key idea is that a page has a high rank if it is pointed to by many highly ranked pages. So, the rank of a page depends upon the ranks of the pages pointing to it. This process is done iteratively until the rank of all pages are determined.

The rank of a page p can be written as:

$$PR(p) = d/n + (1-d)\sum_{(q,p)\epsilon G} \frac{PR(q)}{Outdegree(q)}$$

Here, n is the number of nodes in the graph and OutDegree(q) is the number of hyperlinks on page q. Intuitively, the approach can be viewed as a stochastic analysis of a random walk on the web graph. The first term in the right hand side of the equation is the probability that a random web surfer arrives at a page p by typing the URL or from a bookmark; or may have a particular page as his/her homepage. Here d is the probability that the surfer chooses a URL directly, rather than traversing a link and 1−d is the probability that a person arrives at a page by traversing a link. The second term in the right hand side of the equation is the probability of arriving at a page by traversing a link.

Hubs and Authorities

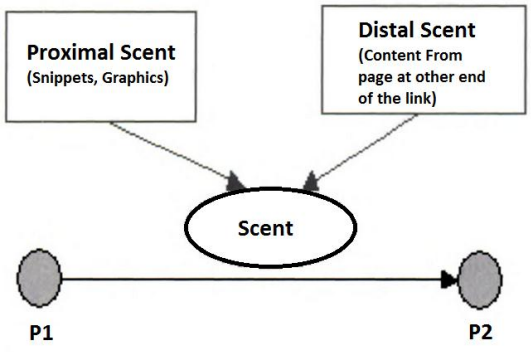
Hyperlink-Induced Topic Search (HITS; also known as hubs and authorities) is a link analysis algorithm that rates Web pages, developed by Jon Kleinberg. The idea behind Hubs and Authorities stemmed from a particular insight into the creation of web pages when the Internet was originally forming; that is, certain web pages, known as hubs, served as large directories that were not actually authoritative in the information that they held, but were used as compilations of a broad catalog of information that led users direct to other authoritative pages. In other words, a good hub represented a page that pointed to many other pages, and a good authority represented a page that was linked by many different hubs.

3.2 Robot Detection and Filtering—Separating Human and Nonhuman Web Behavior

Web robots are software programs that automatically traverse the hyperlink structure of the web to locate and retrieve information. The importance of separating robot behavior from human behavior prior to building user behavior models has been illustrated by Kohavi (2001). First, e-commerce retailers are particularly concerned about the unauthorized deployment of robots for gathering business intelligence at their web sites. Second, web robots tend to consume considerable network bandwidth at the expense of other users. Sessions due to web robots also make it difficult to perform click-stream analysis effectively on the web data. Conventional techniques for detecting web robots are based on identifying the IP address and user agent of the web clients. While these techniques are applicable to many well-known robots, they are not sufficient to detect camouflaged and previously unknown robots. Tan and Kumar (2002) proposed a classification based approach that uses the navigational patterns in click-stream data to determine if it is due to a robot. Experimental results have shown that highly accurate classification models can be built using this approach. Furthermore, these models are able to discover many camouflaged and previously unidentified robots.

3.3 Information Scent—Applying Foraging Theory to Browsing Behavior

Information scent is a concept that uses the snippets of information present around the links in a page as a "scent" to evaluate the quality of content of the page it points to, and the cost of accessing such a page (Chi, Pirolli, Chen, and Pitkow 2001). The key idea is to model a user at a given page as "foraging" for information, and following a link with a stronger "scent." The "scent" of a path depends on how likely it is to lead the user to relevant information, and is determined by a network flow algorithm called spreading activation. The snippets, graphics, and other information around a link are called "proximal cues."



The user's desired information need is expressed as a weighted keyword vector. The similarity between the proximal cues and the user's information need is computed as "proximal scent." With the proximal cues from all the links and the user's information need vector, a "proximal scent matrix" is

generated. Each element in the matrix reflects the extent of similarity between the link's proximal cues and the user's information need. If enough information is not available around the link, a "distal scent" is computed with the information about the link described by the contents of the pages it points to. The proximal scent and the distal scent are then combined to give the scent matrix. The probability that a user would follow a link is then decided by the scent or the value of the element in the scent matrix.

4. POSSIBLE WEBSITE DESIGN QUALITIES TO BE EVALUATED FOR CUSTOMER BROWSING BEHAVIOR

Everyone is talking about good website design, but do we know what it actually means? Do we know how to tell if our own web design is working? Do we know what to look for? Without having a clear metric for measuring the quality, it is hard to recognize the quality. There is more to a successful website than simply looking nice or being able to function. Ideally, a successful website would be built with a specific strategy in mind, it is focused on usability so that visitors can navigate the website with success, it incorporates a style that is pleasing to the user's eye, it is filled with content that is relevant to the users, and it is optimized for search engines. Keeping that in mind, these are a few of the key aspects of a strong website design, complemented with some questions that we may ask ourself when measuring the quality of any website.

4.1 Strategy

A Good Website Layout is backed by Good Strategy. Even the most user-friendly and attractive websites are not successful when they are not achieving the needs of the company. Ask ourself: will our visitors get a clear sense of what we offer and who we are upon arrival at our website? Will our design point our visitors in the direction that we want them to go? Is there a clear strategy that is informing of our design? If the answer to any of these questions are no, then our design is not at its full potential. In order to evaluate the effectiveness of your strategy in the website design, here is a checklist of questions to run through:

• What is the category of the business, and is this obvious on the website?
•What is the purpose of the website and is the design accomplishing this?
• Who is the target audience and how does the design take this into consideration?
•What should the audience do and is this design encouraging them to take that action?

After answering these questions, we should be able to define our brand and then set the specific website goals. This will allow us to align our design according to the goals. When the website is informed by a clear strategy the likeliness that it will succeed raises. We should be able to answer all of the previous questions with confidence that the users will be able to take the action that is intended with very little opposition from the website's overall strategy. The strategy will directly influence the design, as they work hand in hand to ensure that the website's purpose is clear, and there is no confusion about the website. Outline your target audience. Once we know who this is, approach the website as if we are them. Is it obviously apparent that the website can accomplish what the target audience needs it for? If not, take a look back and refine your strategy.

4.2 Usability

This is all about the practical consideration of what makes a good web site design, such as user-friendliness, speed, security, site maps and other technical details, etc. Many of these details are not visually apparent—we will not see a website's security when typing in the URL. Even so, usability will make or break a website. If the visitor is not able to find what they are looking for because of bad navigation, the user will generally leave. If a page takes too long to load, not only will visitors notice but search engines will notice as well. To evaluate how usable our website is ask the following questions:

• How long does it take for pages to load and will the visitors get bored waiting? There are free tools online to test page load speeds.
•Can information be found easily?
•Is there a search button available for visitors?
•Are all the links working? There are tools available online to check a website's links.
•Does the website work in different browsers? Check all of the widely used browsers.
•Does the website work on mobile devices?
•If asking for personal details for taking part in e-commerce, is the customer's information secure? Has this been communicated to the users?

Think of all the ways that will make our website as usable as possible. Imagine that we are coming to it as a visitor and we are trying to find out more information. Additionally, take the extra step in terms of security and be sure to always protect the customer's personal data. The website should be safe to use and should protect the information of the users. If this is not the case, then there will be negative repercussions. Customers will not trust the website at all, and there is a possibility that they will leave negative reviews elsewhere on the internet. The most vital components call-to-action on the website should not be any more than just a few clicks away. If a user has to hunt for the action that they need to complete, they will likely get frustrated and leave the website. They need to be able to navigate smoothly and simply, without the need to guess whether or not they are even on the correct page of the website. Accessibility also falls into the category of usability. It is vital to be sure that your website is accessible to anyone on the internet. This means that it needs to meet or be able to meet accessibility requirements. Anyone with any sort of mental, physical, or cognitive limitations should be able to use the website with little to no alteration. This means that words should be able to be read clearly with the color choices, font choices, and size choices. Creating ALT text on images will allow accessibility software to use the text to audibly describe the images to those with a hearing impairment. In addition to this, the website should be accessible with a keyboard, for those that lack the dexterity in their arms or hands.

4.3 Content and Readability

The two main things to take into consideration when regarding content are its usefulness and readability. Content needs to matter to the readers or we will lose them. Readability is vital, because if the visitors cannot make sense of our content whether it be because it is too small or in a strange color or unreadable font, there's no way to convey the message. These are some questions to ask ourselves when considering the content of the website and how to evaluate its quality:

• Are the fonts that of been chosen easy to read?
•Is there considerable contrast between the font color in the background color?
•Is the text an appropriate size?
•Will the content be relevant to the reader?
•Is the content concise yet still useful?
•Does the overall design make content easy to find?

Evaluate all of the text on the website. Is it conveying the message effectively? Will visitors be able to read the content?

Is what they are reading important to them? Aim for a design that will make all of the content useful and readable. Aside from the aesthetic choice for the content, the quality of the content needs to pull the reader in. Speaking to the visualization that a website is a storefront, you wouldn't walk into a flower shop and find a small selection of flowers, but the walls are stocked with chocolates. This is indicative of a poor, or lack of, strategy. There needs to be some sort of passion behind the strategy to make it clear what exactly the content will be. It is fine to blur the lines of the content every now and again, but the overall message should be clear. Creating content that is consistently relevant will indicate to the visitors of the website that they can grow to expect what type of content can be found on the website, and will increase the chances that they will return to the website.

4.4 Aesthetics

Some say beauty is relative, but that does not mean that there are not defined aesthetic principles that should guide website design. The best type of design will align with the brand, create a positive impression for visitors, be clean, and it will complement the content that is being communicated. To test the effectiveness of the website's aesthetic, ask ourselves the following:

• Does the website's style align with the brand in terms of color, graphics, feel, etc.?

•Is the style consistent throughout the entire website?

• Does the style suit the target audience? (An elegant layout on the website, cartoons on a toy company website etc.)

• How do visitors view the site? Sparse or crowded, orderly or messy, formal or playful? And how does this align with the goals?

•Are there any photos or decorative touches that are getting in the way of the message?

After evaluating these questions, jettison any stylistic choices that directly contradict with the brand's message. Ensure that the logo and website design align. Consider the target audience and let that influence the style. This should not be the most important component (strategy is), but it should go hand in hand with it. The style of the website should complement the brand's goals and intentions, and should never confuse the user of the website. Think of it as the "storefront" of our website. In real life, we'd walk into a professional's office and expect it to be clean, tidy, and quiet. We'd also walk into a party store and expect bright colors, music, and lots of people talking. Imagine if we walked into a party store that was like a professional's office and vice versa. That would leave the customers very confused and they would likely choose not to do business there.

4.5 SEO and Social Networking

There are a lot of ways that the design of a website will impact search optimization. SEO and social networking starts with a strong website design. For example, does the website have a lot of graphics? If it does, remember that the search engines cannot see them. You will need to add ALT tags to the image descriptions so that the search engine will know what is being shown. Is the HTML efficient? If it is not this will hurt search rankings. Consider asking yourself the following questions to ensure that the website design is optimized:

• Are all of the images optimized with ALT tags?

•Is the coding efficient or are there extra lines that can be eliminated?

•Are relevant keywords being used in title tags, heading tags, meta-descriptions, etc.?

•Is there a site map?

A huge mistake is to think that search engine optimization and website design are two separate matters. Consider the way that the design will affect the search rankings and make any adjustments accordingly. The benefits of asking these questions and designing the website accordingly work hand in hand with one another. Not only will adding ALT tags to images aid in the search engine optimization, they will also help in making the website accessible. This is true for all of the heading tags, title tags, and other metadata. A well thought out site map will make navigating the website easy for everyone, especially those that are using the website with the aid of accessibility software.

Based upon these 5 characterizations, how well is any website design working? Think of the ways in which it can be improved. Ponder on the steps that you will take in order to make the design more effective. This could mean improving upon all 5 of the characterizations, or simply improving upon one. It all depends on how well the website was designed in the first place. If anyone wants the website to perform at its absolute best in each aspect. These aspects are all distinctly different, yet they work together to make the website operate smoothly. When all facets of the website operate smoothly, the quality of the website increases, thus bringing more traffic and eventually leads to more conversions, or completes the task intended.

5. CONCLUSION

Web mining is the search for relevant information from the World Wide Web. The found web pages in a search are relevant if they provide an accurate answer to the searcher's information need. As all users of Web search engines are aware, accurate answers are not always at the top of the result list. Web mining is the extraction of interesting and potentially useful patterns and implicit information from artifacts or activity related to the World¬ Wide Web. Before developing a website, one should take in to consideration all the possible quality dimensions, so that the output would be satisfactory.

6. ACKNOWLEDGMENTS

I would like to thank Dr. Shivaji D. Mundhe for guidingme to completethis research paper.

7. REFERENCES

[1] http://dmr.cs.umn.edu/Papers/P2004_4.pdf

[2] https://dynomapper.com/blog/19-ux/188-how-to-evaluate-the-quality-of-your-website-design

[3] https://en.wikipedia.org/wiki/HITS_algorithm

[4] B. Madasamy, Dr. J. JebmalarTamilselvi, "General Web Knowledge Mining Framework", International Journal on Computer Science and Engineering (IJCSE), ISSN : 0975-3397, Vol. 4 No. 10 Oct 2012, 1744-1750

[5] http://www.bea.com/products/weblogic/server/index.shtml

[6] http://www.bvportal.com/

[7] http://www.cio.com/sponsors/110199 vignette story2.html