IJCSIS Vol. 12 No. 12, December 2014
ISSN 1947-5500

International Journal of Computer Science & Information Security


Cornell University
Library


Cogprints


Google scholar


.docstoc
find and share professional documents


ScientificCommons

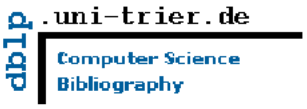

View my documents on
Scribd


BASE
Bielefeld Academic Search Engine


SCIRUS
search engine for science


SciRate.com


CiteSeerx beta


dblp .uni-trier.de
Computer Science
Bibliography


Q·Sensei


DOAJ
DIRECTORY OF
OPEN ACCESS
JOURNALS


EBSCO
HOST


ProQuest

IJCSIS

ISSN (online): 1947-5500

Please consider to contribute to and/or forward to the appropriate groups the following opportunity to submit and publish original scientific results.

CALL FOR PAPERS
International Journal of Computer Science and Information Security (IJCSIS)
January-December 2015 Issues

The topics suggested by this issue can be discussed in term of concepts, surveys, state of the art, research, standards, implementations, running experiments, applications, and industrial case studies. Authors are invited to submit complete unpublished papers, which are not under review in any other conference or journal in the following, but not limited to, topic areas.
See authors guide for manuscript preparation and submission guidelines.

Indexed by Google Scholar, DBLP, CiteSeerX, Directory for Open Access Journal (DOAJ), Bielefeld Academic Search Engine (BASE), SCIRUS, Scopus Database, Cornell University Library, ScientificCommons, ProQuest, EBSCO and more.

Deadline: see web site
Notification: see web site
Revision: see web site
Publication: see web site

Context-aware systems
Networking technologies
Security in network, systems, and applications
Evolutionary computation
Industrial systems
Evolutionary computation
Autonomic and autonomous systems
Bio-technologies
Knowledge data systems
Mobile and distance education
Intelligent techniques, logics and systems
Knowledge processing
Information technologies
Internet and web technologies
Digital information processing
Cognitive science and knowledge
Agent-based systems
Mobility and multimedia systems
Systems performance
Networking and telecommunications
Software development and deployment
Knowledge virtualization
Systems and networks on the chip
Knowledge for global defense
Information Systems [IS]
IPv6 Today - Technology and deployment
Modeling
Software Engineering
Optimization
Complexity
Natural Language Processing
Speech Synthesis
Data Mining

For more topics, please see web site https://sites.google.com/site/ijcsis/

For more information, please visit the journal website (https://sites.google.com/site/ijcsis/)

Editorial
Message from Managing Editor

*The **International Journal of Computer Science and Information Security** (IJCSIS) promotes research publications which offer a significant contribution to the computer science knowledge, and which are of high interest to a wide academic/research/practitioner audience. Coverage extends to all main-stream and state of the art branches of computer science, security and related information technology. As a scholarly open access peer-reviewed journal, IJCSIS mission is to provide an outlet for quality research & academic publications. It aims to promote universal access with equal opportunities for international scientific community; to scientific knowledge, and the creation, and dissemination of scientific and technical information.*

*IJCSIS archives all publications in major academic/scientific databases. Indexed by the following International agencies and institutions: Google Scholar, Bielefeld Academic Search Engine (BASE), CiteSeerX, SCIRUS, Cornell's University Library EI, Scopus, DBLP, DOI, ProQuest, EBSCO. Moreover, Google Scholar reported increased in number cited papers published in IJCSIS (**No. of Cited Papers: 524, No. of Citations: 1106, Years: 5**). Abstracting/indexing/reviewing process, editorial board and other important information are available online on homepage. By supporting the Open Access policy of distribution of published manuscripts, this journal ensures "free availability on the public Internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of [published] articles".*

*IJCSIS editorial board, consisting of international experts, guarantees a rigorous peer-reviewing process. We look forward to your collaboration. For further questions please do not hesitate to contact us at **ijcsiseditor@gmail.com**.*

,

A complete list of journals can be found at:
http://sites.google.com/site/ijcsis/

IJCSIS Vol. 12, No. 12, December 2014 Edition

ISSN 1947-5500 © IJCSIS, USA.

Journal Indexed by (among others):

TABLE OF CONTENTS

Dr. M. Z. Rashad, Dept. of Computer Science, Faculty of Computers and Information Sciences, Mansoura University, Mansoura, Egypt

Abstract - Human intention driven textual analysis is the identification of intentions from natural language text. Human intentions have been widely studied in psychology and behavior sciences, as they are an important feature of human nature, it has also attracted the attention of researchers in computer science especially in the field of human computer interpretation. The ways people use words convey a great deal of information about themselves, their audience, and the situations they are in. Individuals' choice of words can hint at their social status, age, sex, and motives [1]. For more than 50 years, linguists and computer scientists had tried to get computers to understand human language using semantics software. We're still in a long way from having a computer that can understand language as well as a human being does, but we've made definite progress toward that goal [2]. Intention Mining is a new subject and stays at the crossroads of Information Retrieval, Information Extraction and Web Mining. In this survey we presented a basic background that covered the progress of intentions mining field. Furthermore we mentioned the related work in the intention mining and its applications.

Keywords- Human Intentions Knowledge Base, Human Intentions Detection System, and Web Mining.

4. Paper 30111403: Analyzing Different Machine Learning Techniques for Stock Market Prediction (pp. 17-23)

Waqas Ahmad, NUST, College of Electrical and Mechanical engineering Rawalpindi Pakistan

Abstract— Stock market prediction is one of most challenging issue and attracted attention from many researches and stock market investors. With passing time these stock market prediction techniques are getting better with different machine learning algorithms and investors have started relying on these prediction model proposed by many researchers. Many machine learning techniques for stock market prediction are developed. There are no specifications available that which techniques are optimal or not. Also I will analyze and compare different techniques and will discuss their strength and weakness. I have analyzed how these technique works and compare these techniques with other stock market prediction techniques and explained how some techniques have advantage over others and perform better.

Keywords— Stock Market, Machine Learning, Neural Networks, Rough Set, Time Series, Support Vector Machine, Hidden Markov Models, Prediction

5. Paper 30111411: Steganography Using Genetic Encryption Along With Visual Cryptography (pp. 24-31)

Shruti Sekra, Samta Balpande, Karishma Mulani
G. H. Raisoni Institute of Engineering and technology, Wagholi, Pune, Maharashtra, India

Abstract - Authentication login plays a major rule in today's world. Due to unavoidable hacking of the databases, it is always quite difficult to trust the information. The project work aims to solve the problem of authenticity. In this paper, we are proposing a technique utilizing image processing, Steganography and visual cryptography, and then dividing it into shares. In this project the message or the text file is taken as an input from the user which needs to get embedded in the image file. The image file can be of the extensions .jpg or .png. It focuses on hiding secret messages inside a cover medium(image). The most important property of a cover medium is the amount of data that can be stored inside it without changing its noticeable properties. There are many sophisticated techniques with which to hide, analyze, and recover that hidden information. This paper discusses an exploration in the use of Genetic Algorithm operators on the cover medium. Elitism is used for the fitness function. The model presented here is applied on image files, though the idea can also be used on other file types. Our results show this approach satisfied both security and hiding capacity requirements.

Keywords—Genetic Algorithm, Steganography, visual cryptography, Encryption, Decryption

6. Paper 30111413: Performance Evaluation of Statistical Approaches for Automatic Text-Independent Gender Identification System (pp. 32-38)

R. Rajeswara Rao, Department of CSE, JNTUK Kakinada-Vizianagaram, AP, India

Abstract - In this paper, robust feature for Automatic text-independent Gender Identification System has been explored. Through different experimental studies, it is demonstrated that the timing varying speech related information can be effectively captured using Hidden Markov Models (HMMs) than Gaussian Mixture Models (GMMs) . The study on the effect of feature vector size for good Gender Identification demonstrates that, feature vector size in the range of 18-22 can capture Gender related information effectively for a speech signal sampled at 16 kHz, it is established that the proposed Gender Identification system requires significantly less amount of data during both during training as well as in testing. The Gender Identification study using robust features for different states and different mixtures components, training and test duration has been exploited on TIMIT database.

Keywords - Gaussian Mixture Model (GMM),Ergodic Hidden Markov Models (EHMM) Gender, LPC, MFCC.

7. Paper 30111421: Book Related Information Retrieval Using Ontology Based Semantic Search and WordPress (pp. 39-43)

P B Lahoti, Computer Science & Engineering, Government College of Engineering, Aurangabad, India.
V A Chakkarwar, Computer Science & Engineering, Government College of Engineering, Aurangabad, India

Abstract - Nowadays most of the contents are stored and shared on the web, so it is difficult to an intelligent user to find the exact content when it comes to finding and properly managing information in massive volumes. This difficulties are occurred due to Traditional keyword search engine model. Aiming to solve the limitations of keyword-based search engine Semantic Web (SW) has been introduced. Main idea of this research paper is to explore the current state of the semantic information retrieval with major focus on ontology based search. The research paper includes introductory knowledge on the Semantic Web and its layer cake, proposed method which make use of WordPress which is a free and open source blogging tool and a content management system (CMS) based on PHP and MySQL.

Keywords-Semantic retrieval, Ontology based Information search, Book search, WORDPRESS, Ontology based Data Integration (ONDINE), MIEL++.

A Framework for Offline Biometric Identity Authentication

Hanan Elazhary
Computers and Systems Department
Electronics Research Institute
Cairo, Egypt

***Abstract*—Identity authentication based on held paper-based documents such as paper identification cards (IDS) and Hajj permits is not always reliable due to many reasons such as the unclear or relatively outdated photographs in IDs. A better more secure and more reliable approach involves matching biometric features such as fingerprints and iris images. Unfortunately, such an approach typically relies on matching an input live feature (obtained from the stakeholder whose identity is to be authenticated) against a similar feature belonging to the assumed stakeholder but stored in a database. This requires maintaining an expensive local or remote database accessed through a reliable Internet connection and having the access rights to such a database and this is not always feasible. In this paper, we propose a framework for offline biometric identity authentication that does not require the presence of a database (during authentication). It relies on embedding a biometric feature (watermark) in the paper-based document and extracting it at the time of authentication for matching and identity validation. The same framework is suitable for the copyright protection of digital images where the identity of the owner is authenticated offline without having to maintain a database of user-defined watermarks. The framework has different possible implementations and security levels can be added as required with a tradeoff between performance and security. Thus, the paper identifies application requirements that have to be assessed before developing any application-specific implementation of the framework and explains how each of these requirements affect the design and implementation decisions.**

Keywords-Biometrics, copyright protection, identity authentication, image watermarking, security, TA permutation

I. INTRODUCTION

Identity authentication has vast applications in our daily life. For example, it is required at banks and when entering governmental and other different organizations, ports of entry such as airports, schools, and universities. It typically relies on held paper-based documents such as identification cards (IDs) including national IDs and organization IDs (such as university and school IDs). In spite of the widespread use of such paper-based documents, they are not reliable for many reasons. For example, photographs included in these documents are not always clear. People get older and their looks change over time and thus the photographs become outdated. Relatives who resemble each other can also share IDs without being recognized. Hajj permits produced in Saudi Arabia during Hajj are even worse since they are merely printed papers with no included photographs. In an attempt to resolve this issue, many organizations request more than one ID on the same person for identity authentication. But, a better more secure and more reliable approach involves matching biometric features such as fingerprints, iris images, and handwritten signatures possibly in addition to the paper-based documents. Handwritten signatures are usually used at banks and iris images are usually used at airports. The problem with such an approach is that it typically relies on matching an input live biometric feature (obtained from the stakeholder whose identity is to be authenticated) against a similar feature for the assumed stakeholder stored in a database. This requires maintaining a database, which can be very expensive when the number of considered stakeholders is very large. Also, this not always feasible. For example, for the sake of homeland security and privacy concerns, non-governmental organizations may not be authorized to acquire or access such databases. Also, databases are not always stored locally in order to be accessed at different locations and so require the presence of a reliable Internet connection, which is not always available. Even if such connections exist, any failure renders the identity authentication system that relies on them useless. Another problem is that organizations that are inherently visited by the general public cannot rely on databases since possible visitors are unknown in advance. Thus, currently, such an approach has limited applicability.

In an attempt to tackle this problem, this paper proposes a general framework for Offline Biometric Identity Authentication (OBIA). The idea is to embed a biometric feature of a stakeholder in his/her paper-based document used for identity authentication. At the time of authentication, the embedded feature is extracted and matched against a similar live feature obtained from the stakeholder carrying the document to authenticate his/her identity. It is worth noting that the process of embedding an image into another host image/document is termed *watermarking*. The embedded image is the *watermark* and the host image/document is the *watermarked image* [1, 2]. E-documents such as e-passports and smart cards can embed biometric features in a chip within the document, but such documents are expensive and not always available. For example, as mentioned above, Hajj permits produced in Saudi Arabia during Hajj are merely printed papers.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

A similar problem is encountered when attempting to protect the copyrights of digital images by embedding owner-specific watermarks in each image. The typical approach in the literature utilizes any user-specified watermarks, which requires registering to a Certified Authority (CA) [1, 2] that maintains a database of these watermarks. Utilizing owner biometric features as watermarks would allow authenticating the ownership of the images offline without having to maintain and access a local or remote database (during authentication).

This proposed OBIA framework has several advantages: (1) it relies on matching biometric features for increased security and reliability, (2) it is an offline framework that does not require the presence of an Internet connection (to link to a database during authentication) subject to failures, (3) it does not require the presence or maintenance of a large database of biometric features or watermarks for authentication, (4) and so it can be virtually used ubiquitously anywhere and at anytime by any organization, (5) it has several different possible implementations and allows adding levels of security as required according to the application requirements, (6) it is especially useful in case of paper-based identity authentication documents such as Hajj permits, where inspecting the document is combined with matching biometric features for improved security and reliability, and (7) it is also useful for digital image copyright protection since it allows authenticating the identity of the image owner offline. The paper itself has an additional advantage. It identifies application requirements that have to be assessed before developing any application-specific implementation of the framework and explains how each of these requirements affect and guide the design and implementation decisions.

The paper is organized as follows: Section II briefly introduces the image watermarking process and the terminology used in the literature. Section III provides related research in the literature and discusses its shortcomings. Section IV explains the proposed framework, its different modules and its different possible implementations according to the application requirements. Section V presents an example implementation of the proposed framework. Finally, Section VI provides the conclusions and directions for future research.

II. Image Watermarking Overview

Image watermarking refers to the process of embedding a watermark image into a host image [1-2]. Watermarking techniques can be classified in several different ways. According to the visibility of the embedded watermark, watermarking techniques can be classified into:

- Visible watermarking, where the embedded watermark is totally visible such as the case of embedding a logo into the host image.
- Invisible watermarking, where the embedded watermark is totally invisible and have the slightest possible effect on the quality of the host image.
- Transparent watermarking, where the embedded watermark has no effect at all on the quality of the watermarked host image.

According to the watermarking domain, watermarking techniques can be classified into two general classes as follows:

- Spatial domain, where the watermark is embedded into the pixels of the host image.
- Transform domain such as Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT), and Discrete Frequency Transform (DFT).

Watermarking techniques can be also classified according to the robustness of the embedded watermark into three general classes as follows:

- Fragile watermarking, where the embedded watermark is affected by the slightest tampering with the host image.
- Semi-fragile watermarking, where the embedded watermark is affected only by malignant (intentional) transformations, but not by benign (unintentional) ones such as compression for example.
- Robust watermarking that resist unintentional and intentional attacks.

According to the watermark extraction process, watermarking techniques can be classified into three general classes as follows:

- Non-blind watermarking that requires the presence of the original host image and any secret keys used for watermark embedding.
- Semi-blind watermarking that requires the presence of the original watermark image and any secret keys used for watermark embedding.
- Blind watermarking that requires only the secret keys used for watermark embedding, if any.

Finally, according to the host image type, watermarking techniques can be classified into two general classes as follows:

- Color image watermarking techniques
- Grayscale image watermarking techniques

It is worth noting that the watermark image itself can be a binary, grayscale, or color image. But, watermarking techniques are not usually classified according to this criterion.

III. Related Work

Relatively few related research studies exist in the literature. These research studies generally attempt to embed biometric features into host images:

A. Fingerprint

Hasso et al. [3] proposed an algorithm for embedding a grayscale fingerprint (watermark image) into a true color host image. The fingerprint image bits are hidden in the least significant bits (LSBs) of one of the color channels of the host color image without any visual effect on the quality of the watermarked host image. The goal of watermarking was not very clear in the paper. Brindha and Vennila [4] proposed a

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

similar algorithm, but enhanced it in two ways: bits are encrypted before embedding and embedding is performed in a non-linear scattered fashion guided by a pseudo random number generator. The password used for seeding the random number generator is obtained from the user. The goal was to protect biometric data embedded in smart cards. Dutta et al. [5] used Arnold transformation to encrypt the fingerprint watermark and embedded the watermark in the DCT domain of the host color image in an attempt to protect the copyright and ownership of the host image.

B. *Iris image*

Dutta et al. [6] embedded binary code extracted from an iris image using Gabor filter transformation into a color image in the DCT domain for copyright protection. For the same goal, Majumder et al. [7] embedded binary code extracted from an iris image using DCT into a host image in the DWT domain. A pretty similar algorithm was proposed by Lu et al. [8] except that BCH binary code extracted from the iris image using DCT was embedded into the DCT domain of the host image.

C. *Face image*

Inamdar and Rege [9] derived a watermark from a color face image using Principal Component Analysis (PCA). The watermark was then embedded into a grayscale host image using Singular Value Decomposition (SVD) transform. The goal was the copyright protection of the host image. It should be noted that the proposed algorithm is semi-blind requiring the existence of a database of color watermark images to be able to select the face with the closest features to these extracted from the host image.

D. *Voice*

Wang et al. [10] attempted to embed two voice watermarks into a grayscale image: a robust watermark for identity authentication and a fragile one to detect any tampering with the host image.

E. *Handwritten signature*

Bandyopadhyay et al. [11] attempted to embed handwritten signature in two color images sent at different times to be later extracted from the two images using genetic crossover. The goal is to secure the handwritten signature during transmission.

F. *Combined features*

Dutta et al. [12] embedded features from fingerprint images and iris images into images in the DCT domain for copyright protection. Wang et al. [13], on the other hand, embedded features extracted from face images and palm images into e-passport grayscale images for identity authentication.

This discussed related work indicates that the idea of embedding biometric features into host images is promising and starting to gain serious attention from researchers. But, the problem with all these research studies is the lack of a set of clear requirements of each of the corresponding application domains that help in judging and comparing the developed systems in addition to guiding corresponding research studies. Thus, the goal of this paper is to propose a general framework that can be used as a starting point for developing any offline biometric identity authentication system. The framework takes into consideration all possible inputs and outputs, but is general enough to be easily tailored according to the application requirements. Accordingly, the paper identifies different application requirements that have to be assessed before developing any application-specific implementation of the framework and explains how each of these requirements affect the design and implementation decisions.

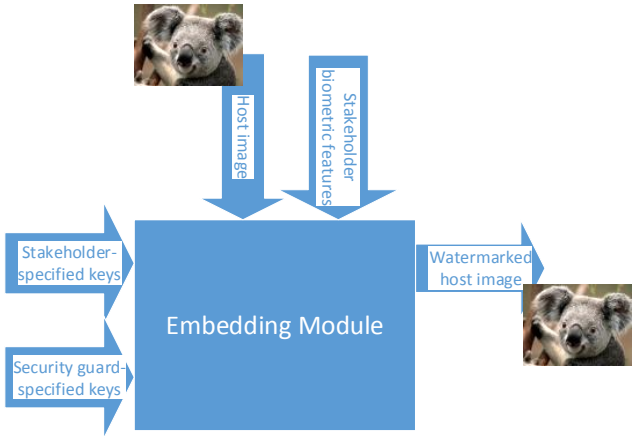


Figure 1. The embedding module.

IV. The Proposed Framework

This section discusses the proposed OBIA framework and its different modules initially in terms of identity authentication based on paper-based documents. We utilize the following terminology: 1) the person/organization responsible for identity authentication is referred to as the *security guard*, 2) the paper-based document is referred to as the *host image*, 3) the biometric features embedded in the host image are referred to as the *watermark*, and 4) the stakeholder whose biometric features are to be embedded in the host image for future identity authentication is referred to as the *stakeholder*.

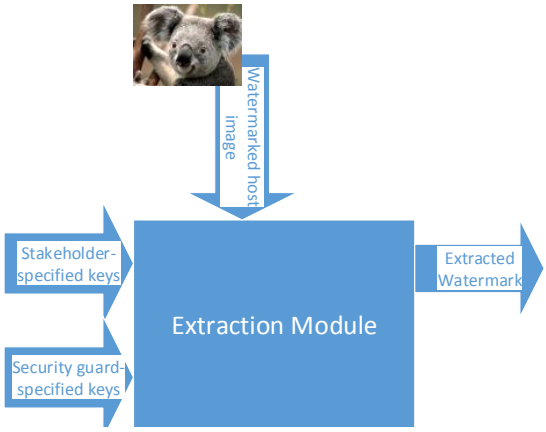


Figure 2. The extraction module.

The framework is composed of three modules: 1) an embedding module for embedding the watermark into the host

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

image, 2) an extraction module for extracting the embedded watermark from the host image, and 3) a matching module for matching the extracted watermark against similar live biometric features obtained from the stakeholder at the time of identity authentication. The embedding module is depicted in Figure 1. The input to this module is the host image and three sets: a set of biometric features of the stakeholder (obtained using relevant biometric scanners), a set of stakeholder-specified keys, and a set of security guard-specified keys. The output of this module is the watermarked host image. This image can be then printed out on the required paper-based document used for identity authentication.

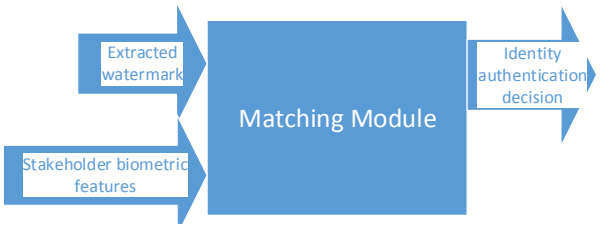


Figure 3. The matching module.

The extraction module, on the other hand, is depicted in Figure 2. The input to this module is the watermarked host image (obtained by scanning the paper-based document) and two sets: the set of stakeholder-specified keys and the set of security guard-specified keys. The output of this module is the watermark extracted from the host image. Finally, the matching module is depicted in Figure 3. The input to this module is the extracted watermark and the live biometric features (obtained from the stakeholder using relevant biometric scanners). The output of the module is a decision about the identity authentication of the stakeholder.

As mentioned before, the same framework is suitable for the copyright protection of digital images. The only difference is that the watermarked image is not printed out on a paper-based document. It is also clear that the OBIA framework takes into consideration all possible inputs and outputs in each module. Thus, it can be used as a starting point for developing any offline biometric identity authentication system. But, in order to decide an application-specific design and implementation of the framework, several application requirements have to be assessed in advance:

A. *Host image type*

It should be decided in advance what type of host image will be utilized in the system since dealing with grayscale images (8 bits) differ from dealing with color images (8 bits or 24 bits) that are formed of three color channels. Besides, if the nature of the application requires for example a color host image, there is no point in implementing a system that considers grayscale host images.

B. *Stakeholder biometric features*

It should be also decided what type of biometric features would be embedded in the host image. It should be noted that biometric features differ in accuracy, cost, and both the complexity and speed of devices required to obtain them [14]. For example, retinal scans and iris images are some of the most unique biometrics while fingerprints and facial images are the most user-friendly [7]. This is in addition to the complexity, speed, and performance of the software algorithms used for embedding, extraction, and matching. Thus, when choosing one or more of these features for embedding, a compromise has to be made between all these factors. Similarly, increasing the number of embedded features increases the security, but has a negative effect on the speed and performance of the identity authentication system.

C. *Stakeholder-specified keys*

Secret keys may be obtained from the stakeholders if they should be involved in the authentication process such as the case of accessing a safe deposit box at a bank. These keys can be used for many purposes such as scrambling/encrypting the watermark before embedding and/or controlling the embedding process such as specifying the embedding locations. It should be noted that this would have a negative effect on the speed of the authentication process and should be taken into consideration in advance.

D. *Security guard-specified keys*

Similar to the stakeholder, the security guard may also provide secret keys to increase the reliability of the identity authentication process. But, similar to the stakeholder case, this would also have a negative effect on the speed of the authentication process and should be taken into consideration during design.

E. *Visibility of the embedded watermark*

One of the factors that have to be taken into consideration is whether the embedded watermark should be visible, invisible, or totally transparent. For privacy concerns, they might be invisible and for accuracy they might be transparent.

F. *Robustness of the embedded watermark*

Most of the image watermarking techniques in the literature are designed to be robust to attacks since they are typically intended for copyright protection of digital images that can be easily attacked and this complicates and slows down the embedding and extraction processes. In case of the OBIA framework, the watermark may be embedded in a paper-based document that is typically secured against attacks. Thus, a compromise should be made between robustness of the embedded watermarks against attacks and the speed of operation. Additional fragile watermarks can be also embedded in case it is required to figure out whether the host image has been tampered with.

G. *The watermarking domain*

The watermark embedded in a transform domain is more robust to attacks than that embedded in the spatial domain. But, the embedding and extraction processes are usually slower. Thus, the decision to work in the spatial domain or one or more transform domains depends on the level of robustness required by the application and whether the host image is subject to

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

attacks (possibly depending on whether the host image is a paper-based document or a digital image).

H. The watermark extraction process

As mentioned in Section II, the watermark extraction algorithm can be blind, semi-blind, or non-blind. In case of the OBIA framework, this should be an offline process that does not require a copy of the original host image or the original watermark. Thus, only stakeholder-specified keys and security guard-specified keys are allowed.

I. The matching process

The matching process is another factor that has to be decided in advance since the choice of an inappropriate algorithm can have a negative effect on the speed and accuracy of the identity authentication process.

J. The sizes of the host image and the watermark image

The sizes of the host image and the watermark image are other factors that have to be taken into consideration since dealing with a small watermark image and a large host image is much easier than dealing with a small watermark image and a small host image as is the case of the photographs embedded in IDs. One reason is that in case of a small host image, we have to be very careful so as not to affect its details.

K. The hardware involved

The hardware involved in the whole process is one of the most critical factors that have to be taken into consideration. This is especially true in case of paper-based documents since the quality of both the watermarked image and the embedded watermark can be greatly affected by printing and scanning. For example, printers usually utilize a dithering process to account for the small number of possible colors in comparison to the true colors of host images.

V. Example Implementation

In this section, we provide an example application with specific requirements. The application-specific implementation details of the OBIA framework according to these requirements are also explained. The assumed requirements are as follows:

- copyright protection of a digital image
- color host image
- grayscale fingerprint watermark
- the host image is of reasonable size with respect to the watermark image
- security of the embedded watermark
- only security guard-specified keys are allowed in the process
- invisible watermark
- fast authentication

To obtain the fingerprint and to perform the matching process, the U.are.U 4500 fingerprint reader [15] has been utilized. This reader has been selected due to its many advantages. For example, it can be interfaced to the computer using a USB and is very fast in scanning a fingerprint and in the matching process too. According to the above requirements, the implementation of the three modules of the OBIA framework is as follows:

A. Embedding module

In this module a color host image is input in addition to a grayscale fingerprint watermark and security-guard specified secret keys. Another key *m* is obtained from the size of the watermark image *m*n*. These keys are used to scramble the fingerprint watermark before embedding using extended Torus Automorphism (TA) permutation. TA permutation can be applied to scramble a watermark image once using *m* and one of the input secret keys *k* as follows:

$$\begin{pmatrix} i^* \\ j^* \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ k & k+1 \end{pmatrix} * \begin{pmatrix} i \\ j \end{pmatrix} \bmod m \quad (1)$$

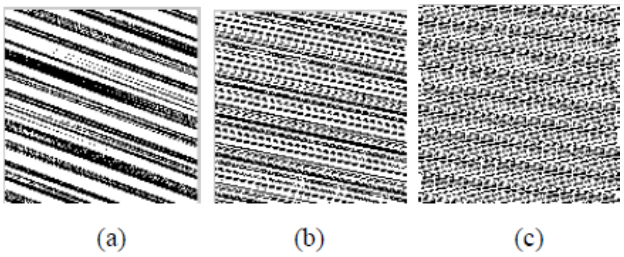


Figure 4. A watermark image scrambled using TA permutation with parameters m = 128, k = 2, and (a) t = 2; (b) t = 4, and (c) t = 8 [2].

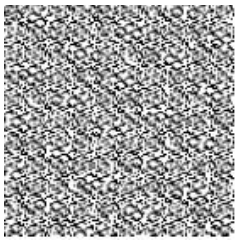


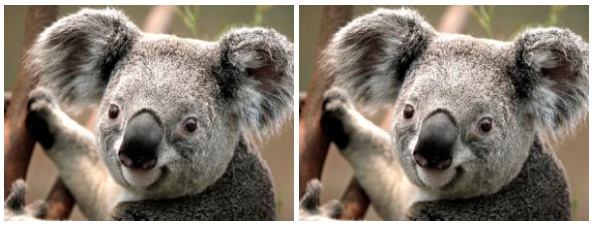
Figure 5. A watermark image scrambled using extended TA permutation with parameters m = 128, t = 3, and k(1)= 2, k(2)=4, and k(3)=8 [2].

In this equation, each pixel (i, j) is moved to a new location (i*, j*). Figure 4 shows a watermark image scrambled using parameters m=128 and k=2 after different number of iterations (t = 2, 4, and 8). To increase the security of the embedded watermark, each of the keys supplied by the security guard is used to scramble the resulting watermark one at a time. Thus, the number of iterations t is equal to the number of supplied keys. This extended version of TA permutation has been proposed in [2]. Figure 5 shows the same watermark scrambled using extended TA permutation using parameters m=128, t=3

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

and a different value for k in each iteration (2, 4, and 8 in iterations 1, 2, and 3 respectively). It should be noted that extended TA has been utilized for scrambling the watermark image since it is highly secure and very fast.

After scrambling the watermark image, it is embedded into the least significant bits (LSBs) of the color channels of the color host image to be totally invisible as specified in the requirements above. Such a simple technique has been selected since according to the requirements above, robustness of the embedded watermark is not an issue and so speed of embedding and extraction has been favored. Figure 6 shows a color host image before and after watermarking. No visual effect is noticed.



(a) (b)

Figure 6. A color host image (a) before watermarking and (b) after watermarking.

B. *Extraction Module*

The input to the extraction module is the watermarked host image and the security-guard specified keys. The scrambled watermark is extracted from the LSBs of the color channels of the host image. The extracted watermark image is then restored using extended torus TA permutation similar to the embedding process, but the k parameters are applied in the reverse order.

C. *Matching Module*

As mentioned above, the U.are.U 4500 fingerprint reader has been utilized in the matching process since it is inherently fast as required in the application. So, no independent algorithm has been developed.

VI. CONCLUSIONS

Most of the current research studies in the literature aim at developing different types of watermarking techniques for different types of images. A major problem with most of these research studies is the absence of clear application requirements before an algorithm is developed. Thus, it is hard to judge or compare existing algorithms. Besides, there are no guidelines for future research studies.

Thus, this paper introduced the OBIA framework that can be used as a first step in developing any offline biometric identity authentication system. It is formed of three modules. The first module is an embedding module for embedding a live biometric watermark (obtained from the stakeholder using a relevant biometric scanner) into a host image. The watermarked host image can be then printed out on a paper-based document used for identity authentication unless it is a digital image whose copyright is to be protected. The second module is an extraction module for extracting the embedded watermark from the watermarked image (may be obtained by scanning the watermarked paper-based document unless the host image is a digital image). The third module is a matching module for matching the extracted watermark against a similar live biometric feature (obtained from the stakeholder using a relevant biometric scanner). This proposed framework has several advantages:

- It relies on matching biometric features for increased security and reliability
- It does not require the presence or maintenance of a large database of biometric features or watermarks for authentication
- It is an offline framework that does not require the presence of an Internet connection (to link to a database during authentication) subject to failures
- It can be virtually used ubiquitously anywhere and at anytime by any organization
- It is especially useful in case of paper-based identity authentication documents such as Hajj permits, where inspecting the document is combined with matching biometric features for improved security and reliability
- It is also useful for digital image copyright protection since it allows authenticating the identity of the image owner offline
- The OBIA framework considers all possible inputs to the three modules, but is general enough to be tailored as required. Different implementations are possible and different levels of security can be added as required according to the application requirements

To wrap up, OBIA framework can be generally used as a starting point for developing any offline biometric identity authentication system. But, researchers should first study the application domain of interest. Requirements of such applications have to be assessed and clearly specified and decided in advance before attempting an application-specific implementation of the OBIA framework. An additional advantage of the paper is that it discusses these requirements and explains the effect of the different requirements on the design and implementation decisions. Thus, the paper can guide future research studies in this area. This would help develop applicable implementations that can be put into practice rather than merely theoretical research studies. Besides, proposed algorithms for a given application can be judged and compared with respect to a clear pre-specified set of application requirements.

An important future research study is the study of suitable algorithms that are unaffected by the printing and scanning of the paper-based documents. Another possible future research study is developing an embedded system that includes for example an inlet for the paper-based document and a scanner for the utilized biometric feature for a faster, easier, and more efficient identity authentication process.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

REFERENCES

[1] H. Elazhary and S. Morkos, "Blind robust transparent DCT-based digital image watermarking for copyright protection," in International Journal of Computer Science and Information Security, vol.8, no. 7, 2010.

[2] H. Elazhary, "A fast, blind, transparent, and robust image watermarking algorithm with extended Torus Automorphism permutation," in International Journal of Computer Applications (IJCA), vol. 32, no. 4, 2011, pp. 34-41.

[3] S. Hasso, M. Hasso, and O. Saad, "Fingerprint hiding in true color image," in International Journal of Computer Science and Information Security, vol.10, no. 8, 2012.

[4] S. Brindha and I. Vennila, "Hiding fngerprint in face using scattered LSB embedding steganographic technique for smart card based authentication system," in International Journal of Computer Applications, vol. 26, no. 10, 2011.

[5] M. Dutta, A. Singh, K. Soni, R. Burget, and K. Riha, "Watermarking of digital media with encrypted biometric features for digital ownership," in Proc. 2013 Sixth International Conference on Contemporary Computing, 2013, pp. 108-112.

[6] M. Dutta, A. Singh, and T. Zia, "An efficient and secure digital image watermarking using features from iris image," in Proc. 2013 International Conference on Control Communication and Computing, 2013, pp. 451-456.

[7] S. Majumder, K. Devi, and S. Sarkar, "Singular value decomposition and wavelet-based iris biometric watermarking," in IET Biometrics, vol. 2. no. 1, 2013, pp. 21-27.

[8] J. Lu, T. Qu, and H. Karimi, "Novel iris biometric watermarking based on singular value decomposition and discrete cosine transform," in Mathematical Problems in Engineering, vol. 2014, article 926170, 2014.

[9] V. Inamdar and P. Rege, "Face features based biometric watermarking of digital image using singular value decomposition for fingerprinting," in International Journal of Security and Its Applications, vol. 6, no. 2, 2012, pp. 47-60.

[10] S. Wang, R. Hua, H. Yua, X. Zheng, and R. Damper, "Augmenting remote multimodal person verification by embedding voice characteristics into face images," in Proc. 2013 IEEE International Conference on Multimedia and Expo Workshops, 2013, pp. 1-6.

[11] S. Bandyopadhyay, D. Bhattacharyya, and P. Das, "Handwritten signature extraction from watermarked images using genetic crossover," in 2007 IEEE International Conference on Multimedia and Ubiquitous Engineering, 2007.

[12] M. Dutta, A. Singh, C. Travieso, and R. Burgee, "Generation of digital signature from multi-feature biometric traits for digital right management control," in Proc. 2014 IEEE Raecs Diet Panjab University Chandigarh, 2014.

[13] Z. Wang, L. Yang, Y. Cheng, and Q. Ding, "Two-stage verification based on watermarking for electronic passport," in Proc. 2013 IEEE Ninth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, 2013.

[14] K. Mali, S. Bhattacharya, "Comparative study of different biometric features," in International Journal of Advanced Research in Computer and Communication Engineering, vol. 2, no. 7, 2013.

[15] The Ultimate USB Fingerprint Reader, Crossmatch, http://www.crossmatch.com/UareU4500Reader/ [accessed December 2014].

AUTHORS PROFILE

Hanan Elazhary received her B.Sc. and M.Sc. in Electronics and Communications Engineering from the Faculty of Engineering, Cairo University. She received her Ph.D. from the University of Connecticut, USA in Computer Science and Engineering. Her research interests include artificial intelligence, multimedia watermarking, software engineering, cloud computing, high-performance computing, and computer networks. Currently, she is an associate researcher professor in the computers and systems department at the Electronics Research Institute, Cairo, Egypt.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Evaluate the quality of service in the WiMax environment using the RED-FEC mechanism

Sami Brini
Faculty of Sciences
University of Dammam
Dammam, KSA

Chokri Idoudi
University of Sousse
ISITCom
Sousse, Tunisia

Abdelbasset Trad
College of Technical (Riadh)
TVTC, *KSA*
Riadh, KSA

Habib Youssef
ISITcom (Sousse)
University of Sousse
Sousse, Tunisia

Abstract: The new information technologies and communication technologies (NTIC) are the basis of the knowledge economy. They allow you to store, process and disseminate an increasing amount of data quickly and cost and are a source of more and more important for productivity gains. Our daily, lives have been improved and work became a lot easier. In this article, we are going to discuss the things where modern technology became useful. This paper study aims to explore the effect of transferring the VoIP network via WiMax. A study of IEEE 802.16 is given by comparing with standards of WiFi (IEEE 802.11, 802.11e ...).
A second part of the paper will be devoted to the evaluation by simulation of the transmission performance of VoIP in a WiMax network. In the third part, we propose an architecture based on the use of two mechanisms for QoS RED (Random Early Detection) and FEC (Forward Error Correction). Based on the results, we propose a new architecture that allows us to keep the level of service quality in good condition.
Keywords: VOIP, WiMax, Audio codecs, FEC, RED

I. INTRODUCTION

Today, wireless technologies [1] are increasingly adopted in our daily lives and in the workplace. The flexibility of these networks seem to attract a large number of people.

In recent years, has revolutionized the WiFi wireless networking, but there is already talk of a new technology: WiMAX (IEEE 802.16).

WiMax is a wireless technology that aims to provide wireless high-speed Internet within several kilometers and is intended primarily for metropolitan area networks "MAN" wireless. In these environments the "voice over IP" (VoIP) [11] presents a more attractive technology having as main goal the reduction of communication costs by transmitting voice and data over a network.

However the conditions of the channel wireless network is highly variable (unpredictable time, losses because to transmission errors on the channel, losses because to congestion) which causes a deterioration in the quality of voice transmitted over such networks, especially in the case of a heavy load on the network by other types of data traffic.

This paper is organized as follows. In a first study of IEEE 802.16 is given by comparing with standards of WiFi (IEEE 802.11, 802.11e ...). The second part of the paper will be devoted to the evaluation by simulation of the transmission performance of VoIP in a WiMax network.

In particular, the work includes analysis of the effect of number of stations transmitting a wireless voice traffic load and other data traffic sent simultaneously with the flow of VoIP quality perceived to be estimated by the MOS (Mean Opinion Score) on the time and the rate of packet loss. The impact of rate audio coding used by the sources will be considered by considering an adaptive approach.

Finally an architecture based on the use of two mechanisms for QoS RED (Random Early Detection) and FEC (Forward Error Correction) is proposed.

In what follows, we present the signaling protocol H.323, then we detail the SIP protocol. After we move to expose our analytical modeling of SIP as well as measures undertaken to evaluate the performance of SIP. Then we illustrate and comment on the different results found. Finally we conclude with a conclusion.

II. VOICE OVER IP

VoIP [4] uses Internet technologies instead of traditional telephone networks to transmit voice signals.

In simple terms, that is a VoIP Internet phone service. This technique is also called IP telephony, Internet telephony, broadband phone and voice over broadband.

While the phone is the crowd favorite but the ability to communicate, via monitor screens anywhere in the world without any financial consideration is also an important criterion for companies. Called his neighbor and called on the other end of the world, the price is still that of a local call. This is obviously the financial aspect is the cause of IP telephony.

Because it is a revolution in prices ahead disproportionately low. IP telephony provides a major economic value for a business:

- Strong reduction of phone bill
- Management of data network (IP network such as the Internet) and the telephone network in a star around a switch by a single crew
- Use a single physical network and thus reducing the cost infrastructure

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

III. PRESENTATION OF THE IEEE 802.16

WiMax described in IEEE 802.16 [5], WiMax is a standard wireless broadband. Operating at 70 Mbit / s, it is planned to connect the access point Wi-Fi1 (Wireless Fidelity) to a fiber optic network, or to relay a shared connection to broadband to multiple users. With a theoretical range of 50 km [5], it should eventually developing metropolitan networks(MAN)1 based on a single access point, as opposed to an architecture based on many Wi-Fi hotspots. WiMax aims to provide wireless high-speed Internet within several kilometers and is intended primarily for metropolitan networks. Indeed, the intended scope of the waves is about 50 km. However, this range is theoretical and the real impact is expected to be more around 8 or 10 km. What remains, however, sufficient to provide connectivity across a city.

A. *The Service With Wimax*

The purpose of the service is to connect the end user to a metropolitan area network so that it can access the Internet. This service is usually performed by the DSLAM or Wi-Fi hotspots. For this, the client must have a WiMAX receiver (a chip or a terminal) and be in the field of action (up to 5 km) of an issuer. Transmission between the client and hotspot WiMAX is said to be "non line of sight" (NLOS).

That is to say that the client is not in sight with the antenna. Indeed, buildings or vegetation found in the cities require the signal to be diverted through the use of OFDM modulation frequency. This is where (the service) that plays the future of mobile WiMAX.

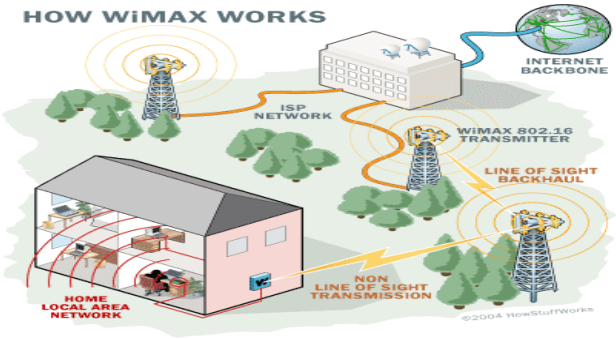


Figure 1. The Service with WiMAX

B. *WiMax As A collector*

In a network, the collection is to connect the access point (Wi-Fi hotspots or DSLAM) to the backbone of the operator (back), thus ensuring the connection to Internet. This is known as the backhaul for hotspots. Unlike the service, the collection is done by "line of sight" (LOS) through WiMAX transmitters placed high enough (of antennas).

At present, this connection is done by wired with fiber optics for example, which is very expensive and cumbersome to implement. The advantage of WiMAX is its simplicity of implementation. It took only two antennas (a few thousand dollars) to connect two remote networks there should have been miles of fiber optic wire.

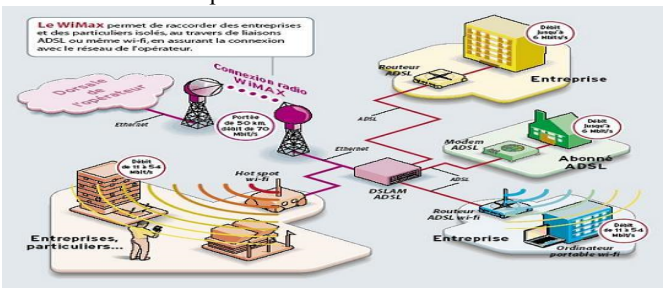


Figure 2. The collection With WiMAX

C. *Transport With Wimax*

Finally, transportation is the step that takes place as "away" from the user. This is to connect the network operator to the global Internet, which involves the use of channel capacity and very high up to distances of several hundred kilometers. WiMAX will have no role to play at this level.

D. *Comparison Between Wi-Fi And WiMax*

IEEE 802.16 [12] networks use the same data link layer (802.2) than other LAN and WAN so they can be bridged and routed them. This is the case for example with the Wi-Fi. The MAC layer in IEEE 802.16 is very different from Wi-Fi in the Wi-Fi, Ethernet uses a contention access method: all users who wish to pass information from one point access to compete for the attention of the access point and get it randomly. The features of WiMAX are better than those of Wi-Fi. WiMAX pushes the limits of Wi-Fi standard by providing increased bandwidth and better encryption. The WiMAX standard also aims to provide increased connectivity between network endpoints without the need for direct eye contact in certain circumstances. Details about the operation without eye contact (Non Line Of Sight - NLOS) are unclear because they have yet to be demonstrated. It is generally considered a spectrum located below the 5-6 GHz is needed to provide a performance of NLOS with reasonable performance and good value for money of PTM (point to multipoint). WiMAX uses the signals to intelligently "routes" but many do not defy the laws of physics [10].

There is no doubt that WiMAX over Wi-Fi performance and yet they will not have the same use. WiMAX and Wi-Fi technologies will coexist and become increasingly complementary to their respective applications. WiMAX is considered as a replacement for the Wi-Fi WiMAX complements Wi-Fi instead of extending its scope. Wi-Fi has been designed and optimized for Local Area Networks (LAN), whereas WiMAX was designed and optimized for Metropolitan Area Networks (MAN).

IV. SIMULATIONS AND RESULTS OBTAINED

In this section we simulate [8] the transfer of voice via the wireless network WiMax and we evaluate the quality of service. To evaluate the performance of VoIP over the WiMax wireless network, we chose several scenarios in order to compare the

[1] : **Wi-Fi** is a technology that can wirelessly connect multiple devices (computer, router, Internet set-top box, etc..) within a computer network.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

results to conclude that helps us with the aim of improve the quality of service of VoIP.

The OPNET environment [6] allows modeling and simulation of communication networks through its model libraries (routers, switches, workstations, servers, WiMax base station ...) and protocols (TCP / IP, FTP, FDDI , Ethernet, ATM, SIP, ...). OPNET is a tool for modeling and simulation of networks is developed and marketed by OPNET Technologies Inc. [OPNET]. It is now a standard reference in the field of network simulation.

A. Restricted Architecture

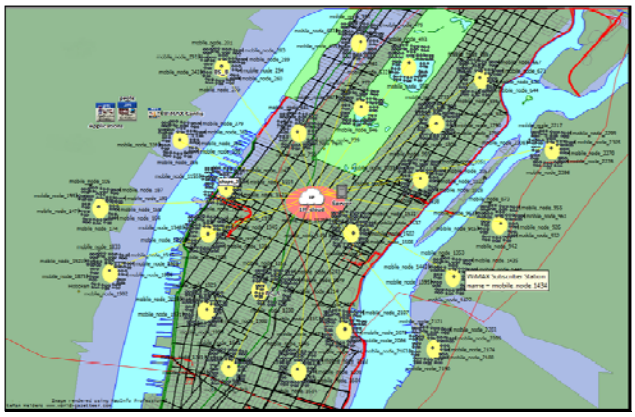


Figure 3. Network Architecture

The overall architecture is the network we want to simulate and examine in order to evaluate quality of service in WiMax. This architecture consists of 23 BS and 6450 mobile clients connected to these base stations. Customers communicate with each other via the VoIP network using WiMax.

Parameter	Attribute	Value
VoIP quality PCM	- Codec - Number of frame per packet - Type of service	* G.711 *1trame/ packet * Interactive Voice
Simulation	- Timing - communication length	* 1 hour * 160 seconds

TABLE I. SOFTWARE CONFIGURATION OF OUR NETWORK

Client Type	number of customer	Parameter	attribute	value
Base station BS	23	WiMax Parameter profile	-Physical Layer - Antenna gain - duplex type - distance between BS and customer - mobile	* OFDMA 5 MHZ * 15 dB * TDD * 2.4 km
Client	6450	WiMax Parameter -Physical Layer - Antenna gain	- Modulation	*64 QAM 3/4 *OFDMA 5 MHZ *-1 dB
VoIP Server	1	Server parameter		Server HP 9000 4 CPU

TABLE II. HARDWARE CONFIGURATION OF OUR NETWORK

B. Results

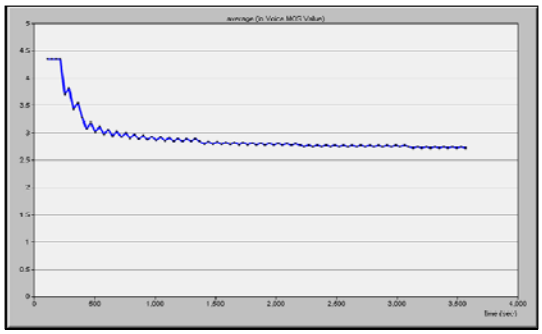


Figure 4. Menas opinion of Score MOS

This figure above represents the measurement of quality of service in our network. If we increase the number of connections, the number of audio streams sent will be increased so the transmission support saturated.

The number of lost packets increases so the quality of service is low. The curve shows the degradation over time, the load increases and the MOS index is 4.4 so the quality is good.after few time the MOS index increased to 2.7 and the quality of service is medium.

C. Results Obtained with the Use of QOS parameters

In this section we build on the results obtained from the simulations to integrate service QOS parameters in order to identify the effects of each parameter on the quality of service received. We use the technique of error correction FEC and management mechanism RED queues in order to obtain a scenario with a stable quality of service with the index 3.5 of MOS. In this context we study the effect of using RED-FEC mechanism on the quality of service and rate of data loss. [2]

1) The RED Mechanism

Random Early Detection (RED) [3] gateways for congestion avoidance in packet switched networks. The gateway detects incipient congestion by computing the average queue size.

The gateway could notify connections of congestion either by dropping packets arriving at the gateway or by setting a bit in packet headers. When the average queue size exceeds a preset threshold, the gateway drops or *marks* each arriving packet with a certain probability, where the exact probability is a function of the average queue size.

RED gateways keep the average queue size low while allowing occasional bursts of packets in the queue. During congestion, the probability that the gateway notifies a particular connection to reduce its window is roughly proportional to that connection's share of the bandwidth through the gateway. RED gateways are designed to accompany a transport-layer congestion control protocol such as TCP.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

The RED gateway has no bias against bursty traffic and avoids the global synchronization of many connections decreasing their window at the same time. Simulations of a TCP/IP network are used to illustrate the performance of RED [2] gateways.

2) The mechanism of error correction FEC

Forward error correction (FEC) [7], error correction technique where there is no transmission of data and therefore the recipient is responsible for correcting errors in the pack. The FEC is based on the sequence of numbers contained in the data field of the ATM protocol AAL (ATM adaptation layer) to detect a loss of cells and avoid unnecessary transmission of cells belonging to the error packets. The FEC technique also allows multiple devices to share the same ATM virtual circuit for transmitting audio and video in real time with minimal overhead (approximately 3%) and slight decrease in performance. The sender adds redundancy to enable the recipient to detect and correct some errors. This prevents the transmission, and thus to save bandwidth, or to ensure the transmission in certain situations where there is no return path.

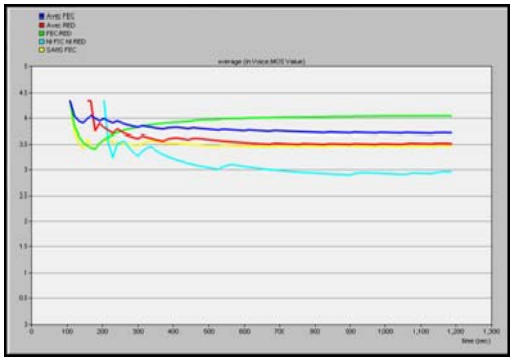


Figure 5. Means opinion of Score MOS Measure

In this figure we present the different scenarios in our study. In the first scenario is presented by the curve in blue sky, not using any QOS parameters, FEC or the RED we note that the index MOS is 3 with a mean quality of service. In the second scenario when using the technique of error correction FEC, MOS is 3.7 that represent a fairly well quality of service. In the last scenario we decide to use both technical and FEC RED same time. The perceived service quality is good and the MOS is 4.

parameter	Attribute	Value
VoIP	- Codec - Number of flows - Type of service	- G.711 Debit : 64 kb/s - 1 frame/ packet - Interactive Voice
simulation	duration	20 min
	Communication duration	160 seconds
Service type	Gold	10 Mb/s
Class service	ertPS	
FEC (Forward Error Correction)	Coding rate	3/4
RED (Random Early Detection)	- Exponential weight factor - Minimum Threshold - Maximum Threshold - Mark probability denominator	9 100 200 10

TABLE III. SOFTWARE CONFIGURATION OF OUR NETWORK

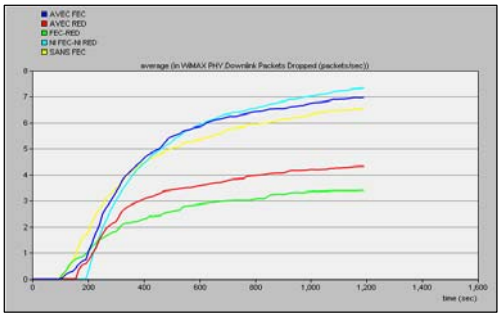


Figure 6. Loss measurement of traffic

In this figure we note that the loss rate is less inferior to the other scenarios if we use the RED-FEC mechanism. The loss rate in our architecture is almost 3 to 4 second per packet but when using one of these two mechanisms the loss rate is increased until it reached 8 packets per second. So the RED-FEC mechanism is less rate loss then the other mechanism. The RED-FEC loss rate is 4% but in the FEC or RED scenario the loss rate is 15%.

V. CONCLUSION

In this paper we study the use of Wimax network to transmit the VoIP flows and the effect in terms of quality of service. We started with a state of the art communication protocol for VOIP and WiMAX networks, we introduced these different specific emphases on the contribution of this relatively new broadband technology in terms desired promotion.

Choosing a wireless technology based on usage that you want to do. WiMax is one of these new technologies. In this context we try in our study using the OPNET simulation software that gives us several analytical curves to simulate several scenarios. These curves allow us to measure voice quality in the environment WiMax. We are changing several parameters such as codec or the QoS parameter.

The simulation results allow us to identify the effects of each parameter or factor on the quality of service in our network.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

The general idea of our approach is the simultaneous use of these two mechanisms RED and FEC. As the case may queue the system decides the number of redundant FEC stream to send this approach allows the reduction of number of lost flow and reduces congestion in the queue. Measurements show that the index is stable MOS perceived value 4 which is a good quality of service. The stability of the MOS index indicates that the mechanism FEC-RED manages the process of transmission of the VoIP WiMax in the environment.

REFERENCES

[1] A. Perrig, R. Szewczyk, V. Wen, D. Culler, and D. Tygar,"SPINS: Security protocols for sensor networks," in Proceedings of Seventh Annual International Conference on Mobile Computing and Networks, July 2001.

[2] A RED-FEC Mechanism for Video Transmission Over WLANs; Cheng-Han Lin, Ce-Kuen Shieh, Naveen K. Chilamkurti, Chih-Heng Ke, and Wen-Shyang Hwang; IEEE TRANSACTIONS ON BROADCASTING, VOL. 54, NO. 3, SEPTEMBER 2008.

[3] A Proposed RED-based Scheduling Scheme for QoS in WiMAX Networks; Po-Chun Ting, Chia-Yu Yu, Naveen Chilamkurti, Wang Tung-Hsien, Ce-Kuen Shieh; Wireless Pervasive Computing, 2009. ISWPC 2009. 4th International Symposium on.

[4] T. Anouari and A. Haqiq, "Analysis of VoIP and Video Traffic over WiMAX Using Different Service Classes", Journal of Mobile Multimedia, Vol. 9, No.3&4, pp. 230-241, 2014.

[5] IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems, IEEE Std. 802.16, 2004.

[6] http//www.opnet.com/solutions/network_rd/modeler.html,

[7] Q. Zhao, C.Li, X.Zhang & D.Liu, "A New FEC Scheme for Real-time Transmission of High Definition Video over IP Network", IPC, 2007, Intelligent Pervasive Computing, International Conference on, Intelligent Pervasive Computing, International Conference on 2007, pp. 232-235.

[8] Z. Abichar, Y. Peng and J. Morris Chang, "WiMax: The Emergence of Wireless Broadband", IT Professional, Vol. 8, Issue. 4, pp. 44-48, Doi:10.1109/MITP.2006.99, July-Aug.2006.

[9] H. Wang, W. Li & D.P. Agrawal. Dynamic admission control and QoS for 802.16 wireless MAN. 2005 Wireless Telecommunications Symposium, pages 60–66, 2005.

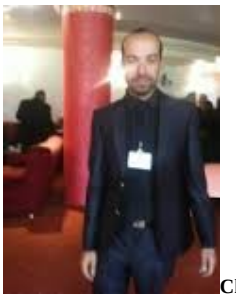
[10] «Comparison of the basic features of Wi-Fi (802.11b/g) - WiMAX (802.16)», Başkent University Department Of Computer Engineering Computer Networks (Bil 431)

[11] M. Adams & M. Kwon, "Vulnerabilities of the Real-Time Transport (RTP) Protocol for Voice over IP (VoIP) Traffic," Proceedings of the 6th IEEE Conference on Consumer Communications and Networking Conference, USA, pp.958-962, 2009.

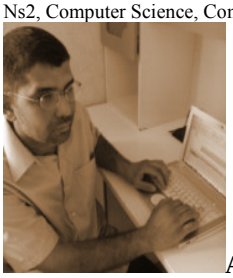
[12] IEEE standard 802. 16-2005, "IEEE standard for Local and Metropolitan Area Networks-Part16: Air Interface for Fixed and Mobile Broadband wireless Access systems Amendment 2", February 28, 2006.

AUTHORS PROFILE

Sami Brini is a PhD candidate at the National School of Computer Science, University Of Manouba-Tunisia. Currently, he is a computer science Lecturer in university of Dammam. His area of research Interest is Wireless Communication, Security and Networking.



Chokri Idoudi received his Master in Information system from the High institute of Data Processing and Management in Kairouan – Tunisia. His research interests include Mobile Adhoc Networks in Ns2, Computer Science, Computer Engineering, and Software Engineering.



Abdelbasset Trad is an Associate Professor at the College of technical - KSA. His area of research Interest is Sensor networks, network security, network protocols, VoIP.

Habib Youssef obtained his PhD in computer science from the University of Minnesota, USA. Currently, he is a Professor at University of Sousse-Tunisia and Director General of Khawarizmi Computing Center (CCK), which is serving as a Data Center and Internet service provider for the Tunisian academic community.

Human Intentions Mining Through Natural Language Text Survey

E. A. Raslan
Bch of Computer Science, Faculty of Computers and Information Sciences, Ain Shams University, Cairo, Egypt

Dr. Y. E. Mohammed
Dept. of Computer Science, Faculty of Computers and Information Sciences, Mansoura University, Mansoura, Egypt

Dr. M. Z. Rashad
Dept. of Computer Science, Faculty of Computers and Information Sciences, Mansoura University, Mansoura, Egypt

ABSTRACT

Human intention driven textual analysis is the identification of intentions from natural language text. Human intentions have been widely studied in psychology and behavior sciences, as they are an important feature of human nature, it has also attracted the attention of researchers in computer science especially in the field of human computer interpretation. The ways people use words convey a great deal of information about themselves, their audience, and the situations they are in. Individuals' choice of words can hint at their social status, age, sex, and motives [1].

For more than 50 years, linguists and computer scientists had tried to get computers to understand human language using semantics software. We're still in a long way from having a computer that can understand language as well as a human being does, but we've made definite progress toward that goal [2].

Intention Mining is a new subject and stays at the crossroads of Information Retrieval, Information Extraction and Web Mining. In this survey we presented a basic background that covered the progress of intentions mining field. Furthermore we mentioned the related work in the intention mining and its applications.

Keywords- Human Intentions Knowledge Base, Human Intentions Detection System, and Web Mining.

1. INTRODUCTION

Understanding intent is an important aspect of communication among people moreover is an essential component of the human cognitive system. The ability to understand the intent of others is critical for the success of communication and collaboration between people, which allows us to "read" others' minds.

The World Wide Web has evolved in less than two decades as the major source of data and information for all domains. Web has become today not only an accessible and searchable information source but also one of the most important communication channels, almost a virtual society.

2. WEB MINING

The web is constantly becoming a central part of social, cultural, political, educational, academic, and commercial life and contains a wide range of information and applications in areas that are of societal interest [6]. The word 'mining' means extracting something useful or valuable, such as mining gold from the earth. The expectation of useful or valuable information discovery from the web is enclosed in the term "web mining". Definitional, web mining refers to the application of data mining techniques to the World Wide Web, or else is the area of data mining that refers to the use of algorithms for extracting patterns from resources distributed in the web. Over the years, web mining has been extended to denote the use of data mining and other similar techniques to discover resources, patterns and knowledge from the web and web-related data [5]. Web mining is a relatively new area, broadly interdisciplinary, attracting researchers from: computer science fields like artificial intelligence, machine learning, databases, and information retrieval specialists; from business studies fields like marketing, administrative and e-commerce specialists; and from social and communication studies fields such as social network analyzers, pedagogical scientists, and political science specialists [4].

3. INTENTIONS MINING

After a person adopts a goal the goal is called an intention and remains an intention until it is met or abandoned, so intention is the thing that you plan to do or achieve. This paper key questions are could we mine the text to understand the inner intentions?, and can we teach computers to reliably and accurately understand human intentions? are of course one of the great challenges of science, and language related technology is one of the great opportunities of information technology due to the need to automatically analyze large amounts of information stored within arbitrary text sources on the internet [1]. Yet, the acquisition of knowledge about common human goals represents a major challenge [3], That is the great range of human Intentions even in relatively restricted domain, which makes the problem really difficult for a computer.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Intentions mining have a variety uses like plan and activity recognition (for language processing and dialogue management, building intelligent user interfaces, etc.), Intelligent personal information management systems, like calendaring agents and to-do list software, Ethnographic analysis of human goals to better understand what goals people have, and Goal-directed inference for natural language understanding. In general, specifying a goal helps a person to achieve it.

4. INTENTION MINING BACKGROUND

Austin (1975), in the theory of speech acts, distinguished between utterances that are statements whose truth or falsity is verifiable and utterances that are not statements. He observed that, "there are, traditionally, besides grammarians statements, also questions and exclamations, and sentences expressing commands or wishes or concessions" [7].

4.1 Discourse Theory

In the introduction to the collection "Intentions in Communication" Cohen et al. (1990) suggest that any theory that purports to explain communication and discourse "will have to place a strong emphasis on issues of intention" [8]. To illustrate the point, they offer a sample dialog between a customer looking for some meat and a butcher selling the same:

Customer: "Where are the chuck steaks you advertised for 88 cents per pound?"

Butcher: "How many do you want?"

The butcher's response would be perfectly natural in a scenario where the steaks are behind the counter where customers are not allowed, and the plausibility of this conversation shows that people infer intention, just as the butcher infers the intention of the customer to be a purchase intention (in this case, possibly as much from the context as from the language). Georgeff et al. (1999) discuss the Belief-Desire-Intention (BDI) Model of Agency based on the work of Bratman (1987) [8].

4.2 Intention Analysis for Sales, Marketing and Customer Service

Aiaioo labs [8], present and attempt to demonstrate the effectiveness of a method of categorization of intentions that is based on the needs of the marketing, sales and service functions of a business which are, according to Smith et al. (2011), the functions most impacted by social media. The categories of intention that they use are purchase, inquire, complain, criticize, praise, direct, quit, compare, wish and sell. They also use another category consisting of sentences that do not express intentions.

4.3 Wishes in Reviews and Discussions

Goldberg et al. (2009) developed a corpus of wishes from a set of New Year's Day wishes and through evaluation of learning algorithms for the domains 'products' and 'politics', showed that even though the content of wishes might be domain-specific, the manner in which wishes are expressed is not entirely so. The definition of the word 'wish' used by Goldberg et al. (2009) is "a desire or hope for something to happen" .

The wish to purchase and the wish to suggest improvements are studied in Ramanand et al. (2010). (Ramanand et al, 2010) propose rules for identifying both kinds of wishes and test the collection of rules using a corpus that includes product reviews, customer surveys and comments from consumer forums. In addition, they evaluate their system on the WISH corpus of (Goldberg et al., 2009). (Wu and He, 2011) also study the wish to suggest and the wish to purchase using variants of Class Sequential Rules (CSRs) [8].

4.4 Requests and Promises in Email

Lampert et al. (2010) study the identification of requests in email messages and obtain an accuracy of 83.76%. A study of email communications by Carvalho and Cohen (2006) and Cohen et al. (2004) focuses on discovering speech acts in email, building upon earlier work on illocutionary speech acts (Searle, 1975; Winograd, 1987) [8].

4.5 Speech Acts in Conversations

Bouchet (2009) describes the construction of a corpus of user requests for assistance, annotated with the illocutionary speech acts assertive, commissive, directive, expressive, declarative, and another category for utterances that cannot be classified into one of those. Ravi and Kim (2007) use rules to identify threads that may have unanswered questions and therefore require instructor attention. In their approach, each message is classified as a question, answer, elaboration and correction [8].

4.6 Sentiment and Emotion

Three of the intentions in the Aiaioo labs study [8], namely the intention to praise something, to criticize something, and to compare something with something else, have been studied by researchers in connection with sentiment analysis.

The detection of comparisons in text has been studied by Jindal and Liu (2006), and the use of comparative sentences in opinion mining has been studied by Ganapathibhotla and Liu (2008). Yang and Ko (2011) proposed a method to automatically identify 7 categories of comparatives in Korean. Li et al. (2010) used a weakly supervised method to identify comparative questions from a large online question archive. Different perspectives might be reflected in contrastive opinions, and these are studied by Fang et al.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

(2012) in the context of political texts using the Cross-Perspective Topic model [8].

The mining of opinion features and the creation of review summaries is studied in Hu and Liu (2006, 2004). A study of sentiment classification is reported in Pang et al. (2002), and the use of subjectivity detection in sentiment classification is reported in Pang and Lee (2004). Studies to detect emotions in internet chat conversations have been described in Wu et al. (2002); Holzman and Pottenger (2003); Shashank and Bhattacharyya (2010). Minato et al. (2008) describe the creation of an emotions corpus in the Japanese language. Vidrascu and Devillers (2005) attempt to detect emotions in speech data from call center recordings [8].

4.7 Analyzing human intentions in natural language text

There is a related work studied human intentions analysis in natural language text that employed the social-psychological theoretical framework [9] that organizes high-level intentions of people into 135 categories. In order to further describe these categories, they attempted to find descriptive phrases by conducting brainstorming sessions. Constructing the Knowledge Base: they identified actions associated with each of the 135 categories by searching for sentences on the web that contained both (i) one of the descriptive phrases for the category, and (ii) an action based causal relation. To build the knowledge base, they constructed a series of query strings by concatenating each descriptive phrase with the following two causal relation phrases: “in order to” and “for the purpose of”. Then, exact phrase searches were issued to the web using the Yahoo! BOSS API. Result page sentences that contained the phrase were stored in an Apache Lucene index. To automatically generate an intent profile, they first segment a given document into a set of sentences. Then, each sentence is issued as a query to the knowledge base. Using the default Lucene similarity measure. They presented a prototypical implementation of an automated method for intent analysis that generates intent profiles of natural language text documents. Their results indicate the potentials of Intent Analysis as a quick, visual evaluation of natural language text from an intentional perspective [10].

4.8 Knowledge Base for Human Intentions

We proposed in this work [12] a technique that build human intentions knowledge base, which has been extracted from 43things Online Social Network as a three level hierarchy and a human detection system with an overall detection accuracy of 76.29%.

Our proposed technique to build Intentions KB shown in figure 1. Firstly, we extracted the 43things data on two stages, the first one to collect the human intentions dataset from 43things, and then we scraped the 43things entries for how to achieve each intention. The next process is cleaning the entries results from unnecessary information, followed by data integration process. Finally, we extract the key features from the cleaned entries results to build our human intentions knowledge base.

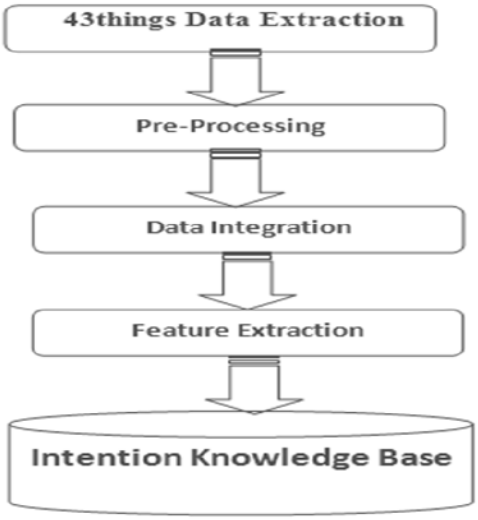


Figure 1. Building Intention Knowledge Base Framework

Our intentions knowledge base presented as a hierarchy of three levels, the top level contains a 47 categories each one of them presented as a vector of 1,000 keywords and their weights, and the sub level contains a 462 categories that contain the 17,615 intentions files and we represented each category of them by a vector of 100 key words and their weights. we determine the Vectors size by building a document-term matrix for each category in each hierarchy level, and take the smallest one of them as the size of the Vector.

We used a hierarchical classification approach that divides a hierarchical problem into a set of flat classification problems, one for each level of the hierarchy. Each class level is treated as an independent classification problem.
In addition, we present results from a study that focused on evaluating intent profiles generated from transcripts of Egyptian presidential candidate speeches in 2014 and American presidential candidate speeches in 2008.

5. Conclusions and future works

We have shown that automatic human intentions detection from a text is possible. We have also presented an algorithm to do so using the massive available data that exists in World Wide Web especially social networks. In addition, we can claim that the ways people use words convey a great deal of information about themselves, their motives and intentions.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

We can compare the result with (M. Kröll, 2009) work, they organize high-level intentions of people into 135 categories only. on the other hand, we build a three levels of a hierarchy taxonomy contains a 17,615 intentions. Moreover, they describe these categories by conducting brainstorming sessions, otherwise, we used a real subscribers entries about how to achieve their goals from 43things social network. However, covering the great range of human intentions even in relatively restricted domain is a challenging problem.

Our ability to distinguish between multiple word meanings is rooted in a lifetime of experience. We can quickly differentiate between the ‘charge’ of a battery and criminal ‘charges’. Using context, an intrinsic understanding of syntax and logic, and a sense of the speaker's intention, we discern what another person is telling us. For computers, this process is not so simple. One of our drawbacks is our dependency on the similarity between intentions titles only in the middle-level, that there is a redundancy in the knowledge base like "be financially independent" , and "be financially responsible". Both of these intentions mean the same. Moreover it causes miss-clustering like "never look back", "look at the sky", and " improve my looks". We suggest to use WordNet lexical database [11] to cluster and classify the intentions.

The Web has become the place for accessing any type of information. There are billions of Web pages and, everyday, new content is produced. regards to this huge data, we suggest to process our Web Mining framework through cloud computing. Cloud computing is clearly one of today's most seductive technology areas due at least in part to its cost efficiency and flexibility. Also we scraped more than 3 million and half WebPages, and only used 35,000 of them because they didn't contain enough entries. We recommend to enrich the knowledge base using other social networks and Wiki sites.

REFERENCES

[1] Pennebaker, J.; Mehl, M. & Niederhoffer, K. (2003), 'Psychological aspects of natural language use: Our words, our selves', Annual review of psychology 54 (1) , 547--577.

[2] "When Will My Computer Understand Me", Linguists, computer scientists use TACC supercomputers to improve natural language processing, https://www.tacc.utexas.edu/news/feature-stories/2013/when-will-my-computer-understand-me., Aaron Dubrow, Science and Technology Writer June 7, 2013.

[3] Shrivastava, V.; Khan, M.; Chaudhari, V.K., "Neural network learning improvement using K-means clustering algorithm to improve the performance of web traffic mining," Electronics Computer Technology (ICECT), 2011 3rd International Conference on , vol.1, no., pp.78,82, 8-10 April 2011.

[4] Pujara, Brinda S., and Sahista S. Machchhar. "Knowledge Discovery from Semantic Web Data using Data Mining Techniques: A Survey." Data Mining and Knowledge Engineering 5, no. 11 (2013): 398-402.

[5] Chang, George, ed. Mining the World Wide Web: an information search approach. Vol. 10. Springer, 2001.

[6] Lappas, Georgios. "An overview of web mining in societal benefit areas." Online Information Review 32, no. 2 (2008): 179-195.

[7] Austin, J. L. (1975). How to do things with words (Vol. 1955). Oxford university press.

[8] Cohan Suja y Carlos, Madhulika Yalamanchi1, Intention Analysis for Sales, Marketing and Customer Service, Aiaioo Labs, India.

[9] Chulef, A. S., Read, S. J., & Walsh, D. A. (2001). A hierarchical taxonomy of human goals. Motivation and Emotion, 25(3), 191-232.

[10] Kröll, M., & Strohmaier, M. (2009, September). Analyzing human intentions in natural language text. In Proceedings of the fifth international conference on Knowledge capture (pp. 197-198). ACM.

[11] http://wordnet.princeton.edu/.

[12] E. A. Raslan, Y. E. Mohammed, M. Z. Rashad, (2014, September). Knowledge Base for Human Intentions: Case Study in 43things Social Network. RECENT SCIENCE PUBLICATIONS, International Journal of Advanced Computing, ISSN: 2051-0845, Vol.47, Issue.2 1522.

Analyzing Different Machine Learning Techniques for Stock Market Prediction

Waqas Ahmad
NUST, College of Electrical and Mechanical engineering Rawalpindi Pakistan

***Abstract*— Stock market prediction is one of most challenging issue and attracted attention from many researches and stock market investors. With passing time these stock market prediction techniques are getting better with different machine learning algorithms and investors have started relying on these prediction model proposed by many researchers. Many machine learning techniques for stock market prediction are developed. There are no specifications available that which techniques are optimal or not. Also I will analyze and compare different techniques and will discuss their strength and weakness. I have analyzed how these technique works and compare these techniques with other stock market prediction techniques and explained how some techniques have advantage over others and perform better.**

***Keywords*— Stock Market, Machine Learning, Neural Networks, Rough Set, Time Series, Support Vector Machine, Hidden Markov Models, Prediction**

1. INTRODUCTION

Since the stock market was established people have earned huge profits but there is huge risk associated with it and there is equal chance of losing money. As result investors have acknowledged need for tools and technique which can help them in stock market prediction. It is one of most challenging issue to correctly predict stock market trends. This has attracted attention of many researchers in mathematics, engineering and finance. Also with the arrival of better computers and easy access of information available on Internet, stock markets data have become easily available to people. Because of this researchers have proposed many techniques for optimal solution in predicting stock.

Many machine learning techniques (MLT) and Data Mining algorithms are developed for stock market. There are no specifications available that which techniques are optimal or not for predicting stock. The main aim of this research is to find out the optimal technique for predicting stock market. Which one is better technique for stock prediction? Can we get a better technique and model by combining multiple predictions? On what basis one prediction system is preferred on other?

2. STOCK MARKET PREDICTION TECHNIQUES

Prediction is to tell what will happen in future on basis of past history and data. Because of non-linear behavior of stock data it is very difficult to predict stock market trends. There are some external factors like Politics, Economy, Terrorism etc. are also involves which create hurdles in stock market prediction. But AI and machine learning techniques have made it possible to predict future market trends to some extent. Few of Stock market techniques are explained below.

2.1 Artificial Neural Networks (ANN)

It is a mathematical model established by W.S.Mcculoch and W. Pitts, and it was name as MP model. It was made by simulating biological nervous systems like the brain. It has following functions:

I. Receive inputs
II. Weight assignment to inputs
III. Calculate weighted sum of inputs
IV. Comparing result with threshold
V. Determine output

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

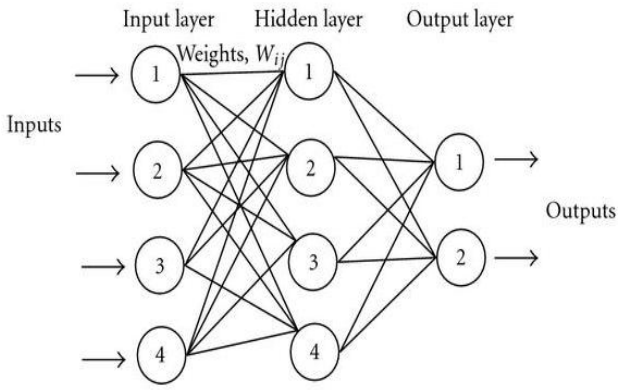


Figure1: Artificial Neural Network (ANN)

It has been claimed in many researches and literature that ANNs are more appropriate for stock prediction than other MLT because ANNs has the ability to find out non linear associations between the training input & output, because of which it makes ANN ideal to produce non linear systems like for stock markets [8].

In ANN we don't assume the functional form of the relationships, it has ability to find out relationship through the data itself. ANNs are known as a universal approximater, when enough data for the modeling are given with the help of ANN any association can be modeled to some level of accuracy. Also, it gives a tolerance to noise and incomplete data. On the other hand ANNs does not show the importance of each of the attribute and how they weigh independent attributes [5].

A very simple approach is shown in [1]. The author used very simple architecture of Artificial Neural Networks. He performed pre-processing on data; he used "Relevance Attribute Analysis" method to remove unwanted attributes and then applied "min-max" normalization. This decreased risk of error.

2.2 Rough Set Model (RSM)

Pawlak introduced Rough Set [7]. It was developed based on mathematical tool, which deals with ambiguity and uncertainty in the classifying the objects in a set. In Rough Set, to organize data decision table are used. This table contains attributes and data elements. In columns we put attributes and in rows we put data element. Analysis of "limit discernibility" is Main idea of Rough set Model. Based on "indiscernibility" redundant features gets identified and eliminated, to lessen the number of features. Three regions got defined by Rough Set. These regions are based on equivalent classes which are induced by attribute values. These regions are

I. Lower approximation
II. Upper approximation
III. Boundary approximation

In [2] author has provided a decision support method for stock traders and analyzes the stocks financial data with RST and applied this technique on Shanghai Stock and grades data in accordance to their importance for the stock market performance. The Main tasks he used

I. Created stock trading data set
II. Performed pre-processing on data
III. Analyzed financial dataset with the data-mining software suite
IV. Produced result & conclusion.

The experimental results & empirical results produced by author indicate that this study gives an easy way for investors in selecting stocks.

2.3 Time Series

Ordered lists of values of one variable/Parameter, that provided in equal time-intervals is called "time series". "Random mathematical statistics theory" is used in analysis of Time series and process to analyze time. this applied at very large in market potential forecasting, control & adjustment, weather hydrology prediction, enterprise operating management, national economy macroeconomic, area complex development plan. It's important source for Estimation and forecast [3]. Continuous of some pattern over time is the prediction like growth in sale, stock market. The common time series methods are

I. ARMA (Auto Regression Moving Average)
II. ARIMA (Auto Regression Integrated Moving Average)

ARMA is integration of AR (Auto Regression) & MA (moving average) models, it is used to predict future values. ARIMA is alteration of ARMA. If mean & variance are constant time series remain stationary else it is non-stationary. These are full of noise that's why non-stationary time series are difficult to predict. Stock prices are non-stationary

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

category and this is reason it is necessary to remove noise first in this technique. [4]

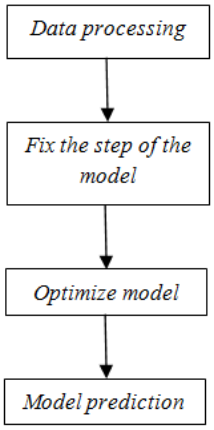


Figure2: ARIMA modeling setup [3]

2.4 Support Vector Machines (SVM)

SVM was developed by Vapnik [6] and rest on "Statistical Learning theory". SVM is a collection of supervised learning techniques that can be applied to regression or classification. SVM has attracted and received attention of many researchers because of its successful applications in regression tasks, classification & financial time-series [9]. SVM has the wanted properties of decision functions control, kernel function use and solutions sparsity [10].

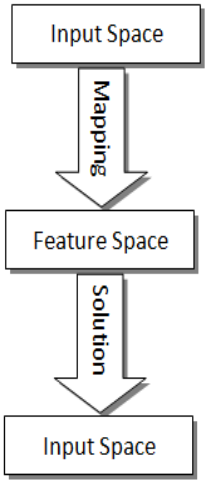


Figure3: SVM Algorithm [11]

It is derived from "Structural Risk Minimization Principle Theory", SVM has displayed characteristic of being very resistant to over-training problem, which as a result achieve low-variance in generalization performance.

SVM results are relatively exclusive and most favorable, not like ANNs training that require non linear optimization with risk to get trapped at local minima.

In [10] author has presented a model or framework for applying Support Vector Machines for stock market prediction. He selected four company-specific & six macroeconomic factors that may influence the stock for further stock multivariate analysis and used SVM to find relationship between these factors and stock performance prediction. It predicts a high percentage of outcomes for many stocks without losing much accuracy.

2.5 Hidden Markov Models

HMM (Hidden Markov Model) can also be used for predicting and forecasting Stock market trends. HMM's have been doing well in predicting and analyzing time-series. In past HMM was used in ECG analysis, speech recognition etc. HMM are based on a collection of unobserved states wherein transitions take place and every state is linked with a collection of possible observations. Stock trends can also be observed in similar way. The underlying states are normally invisible to the investor which determines the behavior of the stock value. The transitions among these states are founded on decisions, economic circumstances & company policies etc. The visible effect which shows these is the stock value. HMM perform well to this real-life circumstances [16].

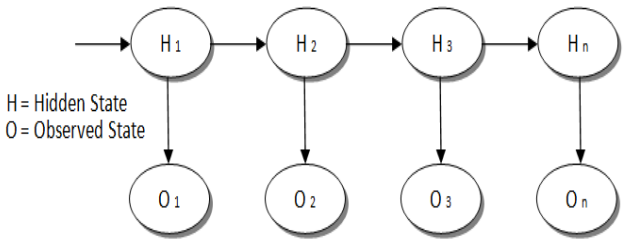


Figure 4: Hidden Markov Model

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

3. HYBRID STOCK MARKET PREDICTION TECHNIQUES

Researchers have combined and mutate different Machine Learning techniques to get more optimal solution for stock market prediction in last few years. I have analyzed few hybrid techniques used for stock market prediction.

3.1 Adaptive Neural Network Model (ANNM)

ANNM is a "Feed-Forward NN". ANNM has a new activation function named "Neuron Adaptive". For the validation of new ANNM researchers have experimented with approximation function and analyzed stock market. ANNM showed many benefits over simple neuron-fixed feed-forward networks for instance faster learning, minimized network-size & encouraging financial analysis. In ANNM researchers has used ANN with Neuron-adaptive Activation Function to simulate the stock market data. This new technique reduced network-size increased training pace & simulation error [12].

3.2 Integrating Neural Network and Rough set

Another technique which is used to predict stock market is by integrating Neural Network and rough set. Researchers have integrated ANN and rough set to predict best possible buy and sell of a share in stock market and they used "Confusion matrix" to assess performance of predicted & observed classes for this models [13]. Their results showed that this model had higher accuracy than the RS model and the ANN model.

Similarly Researchers have combined RNN (Regularized NN) & RS. RS can take out Rule-Knowledge from trained NN, it is used to predict the time-series performance (Stock market). This hybrid model combines the Rule-Reduction capability of RS and high generalization faculty of RNN and showed the effectiveness of this model in stock market prediction [14].

3.3 Integrating Decision Tree (IDT) and Rough set (RS)

In this technique researchers have IDT with RS for predicting stock market. Features are extracted from the stock market past data in this technique, researchers used technical indicators. Then to select the relevant features they used C4.5 decision tree and RS based system for induce rules from these extracted features. After comparing this model performance with a ANN based stock forecasting & a Naïve Bayes based stock forecasting system, this model outperforms the NN based systems and Naive-Bayes stock market prediction algorithms [15].

3.4 Integrating Genetic Algorithm and ANN techniques

Koza introduced Genetic Programming (GP) by developing symbolic regression. It is a computational optimization tool which is used to derive best possible model from time series data. Mutation, crossover and Reproduction are main process of GP. Main factor is the Fitness Function on which final population are based upon. In a population switching nodes is called cross over. Based on fitness function Genetic Programming reproduces for making new generation. Getting and substituting information of one node with those individuals is mutation. Fitness function is used to evaluate new generation (Langdon and Poli 2002)[19].

Two types of NN are combined with GA. GA's identifies input attributes and weights for these attributes [17][18]. ATNN (Adaptive time delay NN) and TDNN (Time delay NN) are used for their capability for saving temporal patterns. GA ATNN & GA TDNN are suggested by [17] for forecasting stock. It is observed from result given by author that, GA ATNN & GA TDNN outperforms individual ATNN, RNN and TDNN [19].

3.5 Integrating Genetic Fuzzy Algorithm and ANN

This hybrid model is developed by integrating ANN, fuzzy logic (FL) and genetic programming (GP). These different techniques are combined to get a more optimal solution for stock market prediction. There are 3 main phases of this model [19][20]:

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

I. Variable selection. SRA (Stepwise Regression Analysis) is applied to select key variable.
II. Divide data, SOM (self-organization map) NN is used for this phase. By dividing data into useful sub components, SOM minimizes the complexity of data.
III. Construct GFS for stock price prediction

SRA is recursive function and search out the independent factors set, variables got enters or removed on each iteration from model [19]. Related data is being combined by SOM. After that genetic-fuzzy system is constructed. To make KB (Knowledge Base) of fuzzy rule based system 2 steps are involved.

I. Evolve rules using GA
II. Tuning fuzzy system database.

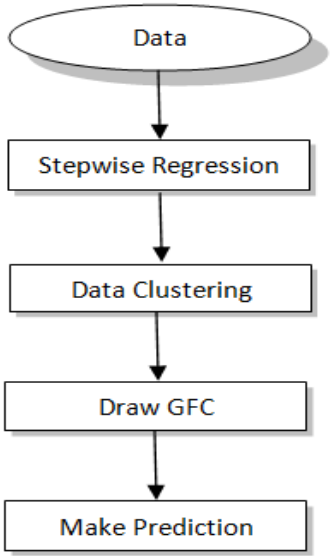


Figure 5: Combining GA with ANN

4. METHODOLOGY

In this paper I have analyzed different machine learning techniques that are being used for stock market prediction. I have reviewed few individual techniques including

I. Artificial Neural Networks (ANN)
II. Rough Set Model (RSM)
III. Time Series
IV. Support Vector Machine
V. Hidden Markov Models

Also hybrid stock market prediction techniques including

I. Adaptive Neural Network Model
II. Integrating Neural Network and Rough set
III. Integrating Decision Tree and Rough set
IV. Integrating Genetic Algorithm and ANN techniques
V. Integrating Genetic Fuzzy Algorithm and ANN

I have analyzed how these technique works and compare these techniques with other stock market prediction techniques and how some techniques have advantage over others and which techniques perform better than other.

In next step I will collect stock data from sources including **OGDCL** Pakistan (Oil and Gas Development Company Limited) and finance.google.com. After collecting data from above mentioned sources I will apply these stock market prediction techniques OGDCL dataset using **MATLAB**. I will use different set of OGDCL dataset, Last one month, Last 3 month, Last 6 Month and finally Last year's complete stock data, to check performance of these techniques that how correctly these stock prediction techniques behave when amount of dataset increases. Also I will compare result of these techniques on basis of time taken and performance measures, that which technique performed better than others on above data sets of OGDCL.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

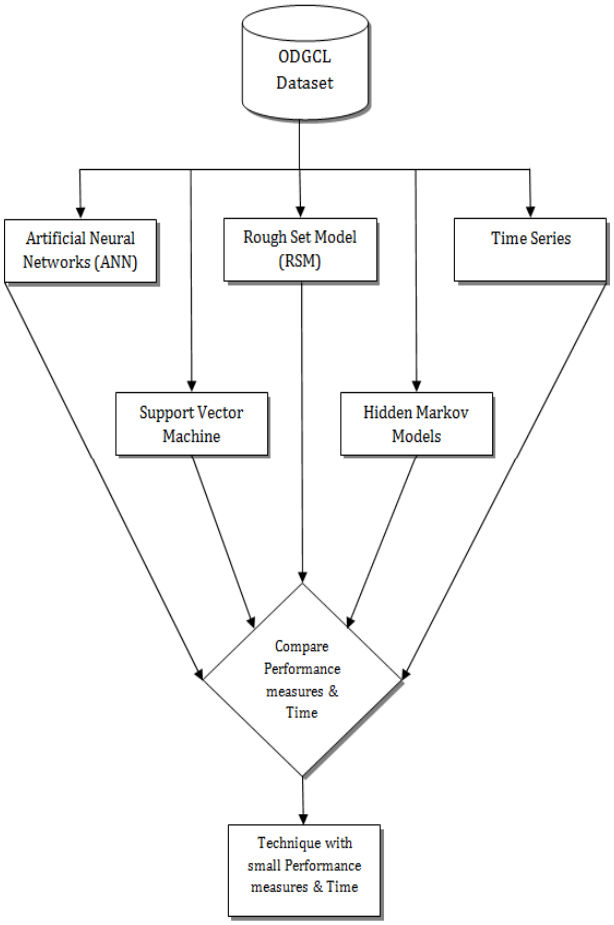


Figure 6: Methodology

5. CONCLUSION

Many machine learning techniques are developed by researchers and are being used for predicting stock-market. In this paper, I have explained few techniques for predicting stock-market by analyzing and comparing different techniques. It was observed that although many state of the art stock market prediction techniques are available but in some techniques it is necessary that we apply data pre-processing & post-processing to achieve better results in prediction. Also it was observed that we can have many combinations of different Machine learning algorithms and we can integrate different techniques to develop new stock market prediction models and produce optimal results.

6. REFERENCES

[1] Parkash Ramani , Dr. P.D.Murarka, 2013, 'Stock Market Prediction Using Artificial Neural Network. International Journal of Advanced Research in Computer Science and Software Engineering, Volume (3)

[2] Wang Ruizhong, 2012 'Analyses the Financial Data of Stocks Based Rough Set Theory'. 2012 Eighth International Conference on Computational Intelligence and Security

[3] Jun Zhang, Rui Shan & Wenfang Su, 2009, 'Applying Time Series Analysis Builds Stock Price Forecast Model', 2009 Journal of modern applied sciences, Vol. 3 , No. 5

[4] Asim Maqbool, Umer Malik, Mohsin Fazal, Mutte Ahmed(2005), 'Stock Market Prediction Software using Recurrent Neural Networks'

[5] Li, F., Sheng, H. and Zhang, D. 2002, 'Event Pattern Discovery from the Stock Market Bulletin', Discovery Science: 5th Int. Conference, vol. 2534, pp. 310-315.

[6] Vapnik, V. 1999, 'An overview of statistical learning theory'. IEEE Transactions of Neural Networks, vol. 10, pp. 988-999.

[7] Z. Pawlak, "Rough sets", International Journal of Computer and Information Science, pp.341-356, 1982.

[8] Li, F., Sheng, H. and Zhang, D. 2002, 'Event Pattern Discovery from the Stock Market Bulletin', Discovery Science: 5th Int. Conference, vol. 2534, pp. 310-315.

[9] Yang, H. Chan, L. and King, I. 2002, 'Support Vector Machine Regression for Volatile Stock Market Prediction', Lecture Notes in Computer Science, vol. 2412, pp. 391.

[10] Zhen Hu , Jie Zhu , and Ken Tse, 'Stocks Market Prediction Using Support Vector Machine'. 2013 6th International Conference on Information Management, Innovation Management and Industrial Engine

[11] A.K. Verma, T. N singh, Sachin Maheshwar,'Comparative Study of Intelligent Prediction Models for Pressure Wave Velocity', Journal of Geosciences and Geomatics. 2014, 2(3)

[12] Dr Shuxiang Xu, Prof Ming Zhang, 'Data Mining - An Adaptive, Neural Network Model for Financial Analysis'

[13] Shipra Banik, A.F.M. Khodadad Khan and Mohammad Anwer, 'Dhaka Stock Market Timing Decisions by Hybrid Machine Learning Technique'

[14] XIAO-YE WANG, ZHENG-OU WANG, 'stock market time series data mining based on regularized neural network and rough set', Proceedings of the First International Conference on Machine Learning and Cybernetics, Beijing, 4-5 November 2002

[15] Binoy.B.Nair, V.P Mohandas, N. R. Sakthivel, 'A Decision Tree- Rough Set Hybrid System for Stock Market Trend Prediction', International Journal of Computer Applications (0975 – 8887) Volume 6– No.9, September 2010

[16] Aditya Gupta, 'Stock Market Prediction Using Hidden Markov Models'

[17] Hyun-jung Kim, Kyung-shik Shin(2007), "Applied Soft Computing", "A hybrid approach based on neural networks and

genetic algorithms for detecting temporal patterns in stock markets", 7 (2007) 569–576

[18] Safdar Ali, Farrukh Aslam. Khan and Waseem Shahzad, "Intrusion detection using bi-classifier-based genetic algorithm, "ICIC Express Letters, Vol. 7, No. 6(A), pp. 2147-2153, December 2010. 1881-803x.

[19] Zahid Iqbal, R. Ilyas, W. Shahzad, Z. Mahmood and J. Anjum, 'Efficient Machine Learning Techniques for Stock Market Prediction', Int. Journal of Engineering Research and Applications Vol. 3, Issue 6, Nov-Dec 2013, pp.855-867

[20] Esmaeil Hadavandi, Hassan Shavandi, Arash Ghanbari (2010) 'Integration of genetic fuzzy systems and artificial neural networks for stockprice forecasting, Knowledge Based Systems'

Steganography Using Genetic Encryption Along With Visual Cryptography

Shruti Sekra[1], Samta Balpande[2], Karishma Mulani[3]
G. H. Raisoni Institute of Engineering and technology, Wagholi, Pune, Maharashtra, India

ABSTRACT

Authentication login plays a major rule in today's world. Due to unavoidable hacking of the databases, it is always quite difficult to trust the information. The project work aims to solve the problem of authenticity. In this paper, we are proposing a technique utilizing image processing, Steganography and visual cryptography, and then dividing it into shares. In this project the message or the text file is taken as an input from the user which needs to get embedded in the image file. The image file can be of the extensions .jpg or .png. It focuses on hiding secret messages inside a cover medium(image). The most important property of a cover medium is the amount of data that can be stored inside it without changing its noticeable properties. There are many sophisticated techniques with which to hide, analyze, and recover that hidden information. This paper discusses an exploration in the use of Genetic Algorithm operators on the cover medium. Elitism is used for the fitness function. The model presented here is applied on image files, though the idea can also be used on other file types. Our results show this approach satisfied both security and hiding capacity requirements.

Keywords—Genetic Algirithm,Steganography, visualcryptography,Encryption,Decryption

I.INTRODUCTION

Steganography is a branch of information hiding. It embeds the secret message in the cover media (e.g. image, audio, video, etc.) to hide the existence of the message. Steganography is often used in secrete communication. In recent years, many successful steganography methods have been proposed. Among all the methods, LSB replacing method is widely used due to its simplicity and large capacity.The majority of LSB steganography algorithms embed messages in spatial domain, such as BPCS, PVD. In the LSB steganography, secret message is converted into binary string. Then the least signicant bit-plane is replaced by the binary string. The LSB embedding achieves good balance between the payload capacity and visual quality. However, the LSB replacing method ips one half of the least-signi_cant bits. Thus the artifacts in the statistics of the image are easy to be detected.

The basic structure of Steganography is made up of three components:

i. The Carrier image,
ii. The Message,
iii. The Key

The carrier can be a painting, or a digital image. It is the object that will „carry" the hidden message. A key is used to decode/decipher/discover the hidden message. This can be anything from a password, a pattern, a black-light etc.

Steganalysis is the method to reveal the hidden messages, even some doubtful media. The attacks on LSB replacing

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

methods are most based on Chi-square analysis and the relationship of pixels or bit planes.

The genetic algorithm is used to estimate the best adjusting mode. By the adjustment, the artifacts caused by the steganography can be eliminated and the image quality will not be degraded. Experimental results of another RS-resistent method are compared with the proposed one, and it is revealed that the proposed algorithm exhibits excellent security and image quality.

II. LITERATURE SURVEY

The simplest insertion method in steganography is LSB replacement steganography. In the LSB replacement method, the least significant bit of the pixel values are replaced with the bit values of the message. The method of detecting the secret message hidden in the cover media through steganography is known as steganalysis. Steganalysis methods are of two types, one that attacks only color images or grayscale images and the other which attacks on both color and grayscale images. However, irrespective of the mentioned type of image, some of the steganalysis methods attack only on LSB embedding, while others attack on different methods which also include LSB embedding. Few of the steganalysis methods suspect the message hidden in the image whereas few other steganalysis methods detect the length of the message hidden in the image.

Arezoo Yadollahpour and Hossein Miar Naimi proposed a steganalysis technique using autocorrelation coefficients in colour and grayscale images. They suggest that insertion of secret message weakens the correlation between the neighbour pixels and thereby enable one to detect the message.

Fridrich et al proposed an effective steganalysis technique popularly known as RS steganalysis, which is reliable even in the detection of non-sequential LSB embedding in digital images.

Andrew D Ker has proposed a general framework for structural steganalysis of LSB replacement for detection and length estimation of the hidden message. He suggests the use of previously known structural detectors and recommended a powerful detection algorithm for the aforementioned purpose.

Tao Zhang and Xijian Ping have proposed a steganalysis method for detection of LSB steganography in natural images based on different histogram. This method ensures reliable detection of steganography and estimate the inserted message rate. However, this method is not effective for low insertion rates.

III.PROPOSED SYSTEM:

The proposed system makes use of both stegnographic as well as visual cryptographic technique.Stegnography uses Genetic algorithm for providing security and the second protection lock used is visual cryptography.So combining both stegnography and visual cryptographic algorithm enhace double security to the system.

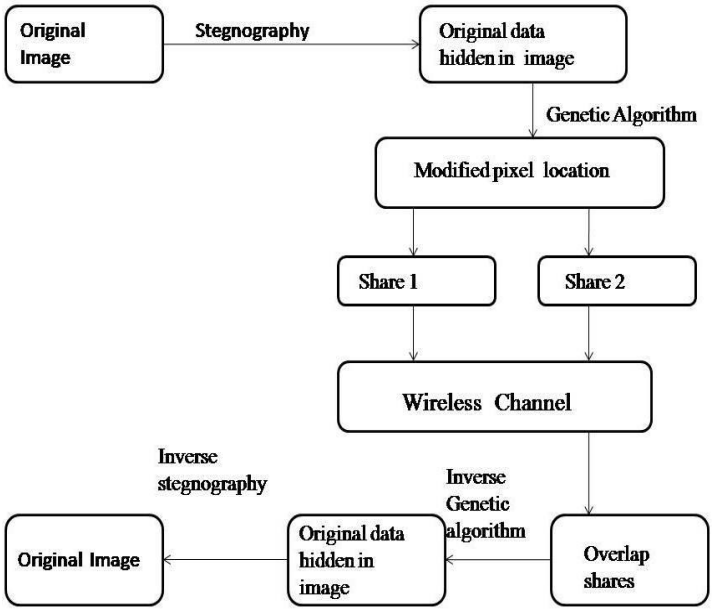


Fig.Proposed Model

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

1.Genetic Algorithm:

Genetic Algorithm (GA) is based on biological evolutionary theories and is often used to solve optimization problems. GA comprises of a set of individual elements (the population) and a set of biologically inspired operators. According to evolutionary theories, only the most suited elements in a population are likely to survive, generate offspring, and transmit their biological heredity to the new generations. GA's are much superior to conventional search and optimization techniques in high dimensional problem spaces due their inherent parallelism and directed stochastic search implemented by recombination operators.

In a genetic algorithm, a population of candidate solutions (called individuals, creatures, or phenotypes) to an optimization problem is evolved toward better solutions. Each candidate solution has a set of properties (its chromosomes or genotype) which can be mutated and altered; traditionally, solutions are represented in binary as strings of 0s and 1s, but other encodings are also possible. A part of the chromosomes is called a gene

Outline of the Basic Genetic Algorithm :

1. [Start] Generate random population of n chromosomes (suitable solutions for the problem)
2. [Fitness] Evaluate the fitness f(x) of each chromosome x in the population
3. [New population] Create a new population by repeating following steps until the new population is complete
    (a) [Selection] Select two parent chromosomes from a population according to their fitness (the better fitness, the bigger chance to be selected)
    (b) [Crossover] With a crossover probability cross over the parents to form a new offspring (children). If no crossover was performed, offspring is an exact copy of parents.
    (c) [Mutation] With a mutation probability mutate new offspring at each locus (position in chromosome).
4. [Accepting] Place new offspring in a new population
5. [Replace] Use new generated population for a further run of algorithm
6. [Test] If the end condition is satisfied, stop, and return the best solution in current population
7. [Loop] Go to step 2

Why Genetic Algorithms?

It is better than conventional AI in that it is more robust. Unlike older AI systems, they do not break easily even if the inputs changed slightly, or in the presence of reasonable noise. Also, in searching a large state-space, multi-modal state-space, or n-dimensional surface, a genetic algorithm may offer significant benefits over more typical search of optimization techniques.

1. Selection Operator
    - key idea: give prefrence to better individuals, allowing them to pass on their genes to the next generation.
    - The goodness of each individual depends on its fitness.
    - Fitness may be determined by an objective function or by a subjective judgement.

2. Crossover Operator
    - Prime distinguished factor of GA from other optimization techniques
    - Two individuals are chosen from the population using the selection operator
    - A crossover site along the bit strings is randomly chosen

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

- The values of the two strings are exchanged up to this point
- If S1=000000 and s2=111111 and the crossover point is 2 then S1'=110000 and s2'=001111
- The two new offspring created from this mating are put into the next generation of the population
- By recombining portions of good individuals, this process is likely to create even better individuals .

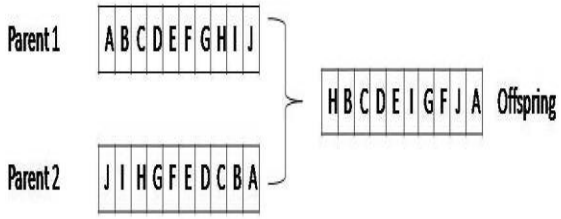


Fig.crossover operation

3. Mutation Operator

- With some low probability, a portion of the new individuals will have some of their bits flipped.
- Its purpose is to maintain diversity within the population and inhibit premature convergence.
- Mutation alone induces a random walk through the search space
- Mutation and selection (without crossover) create a parallel, noise-tolerant, hill-climbing algorithms .

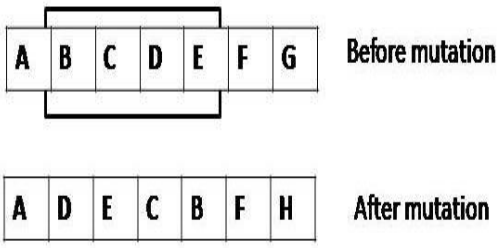


Fig.Mutation operation

2.LSB Algorithm:

LSB (Least Significant Bit) substitution is the process of adjusting the least significant bit pixels of the carrier image. It is a simple approach for embedding message into the image. The Least Significant Bit insertion varies according to number of bits in an image. For an 8 bit image, the least significant bit i.e., the 8th bit of each byte of the image is changed to the bit of secret message. For 24 bit image, the colours of each component like RGB (red, green and blue) are changed. LSB is effective in using BMP images since the compression in BMP is lossless. But for hiding the secret message inside an image of BMP file using LSB algorithm it requires a large image which is used as a cover. LSB substitution is also possible for GIF formats, but the problem with the GIF image is whenever the least significant bit is changed the whole colour palette will be changed. The problem can be avoided by only using the gray scale GIF images since the gray scale image contains 256 shades and the changes will be done gradually so that it will be very hard to detect. For JPEG, the direct substitution of steganographic techniques is not possible since it will use lossy compression. So it uses LSB substitution for embedding the data into images. There are many approaches available for hiding the data within an image: one of the simple least significant bit submission approaches is, Optimum Pixel Adjustment Procedure". The simple algorithm for OPA explains the procedure of hiding the sample text in an image.

Step1: A few least significant bits (LSB) are substituted with in data to be hidden.

Step2: The pixels are arranged in a manner of placing the hidden bits before the
pixel of each cover image to minimize the errors.

Step3: Let n LSBs be substituted in each pixel.

Step4: Let d= decimal value of the pixel after the substitution.

d1 = decimal value of last n bits of the pixel.

d2 = decimal value of n bits hidden in that pixel.

Step5: If (d1~d2)<=(2^n)/2
then no adjustment is made in that pixel.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Else
Step6: If(d1<d2)
d = d – 2^n.
If(d1>d2)
d = d + 2^n.
This„d‟ is converted to binary and written back to pixel
This method of substitution is simple and easy to retrieve the data and the image quality better so that it provides good security.

3.Encryption And Decryption Algorithm:

The different symmetric encryption algorithms are

1. Data encryption standard
2. Advanced encryption standard

1.Data encryption standard (DES):

Data Encryption Standard‟ (DES) is also known as Data Encryption Algorithm (DEA). DEA takes 64 bits of plain text and 56 bits of key to produce 64 bits cipher text block. The DES algorithm always functions on blocks of equal size and uses the permutations and substitutions in algorithm.
The data encryption algorithm uses 56 bit key so it is not possible for the defender for analysing the key. So, the problem of Cryptanalysis is avoided using this algorithm. But the drawback of the algorithm is Brute-force attack. This can be avoided using the Triple DES algorithm.
Triple DES:

Triple DES is an extension to the DES algorithm. Triple DES uses the same approach for encryption as DES. 3DES takes three 64 bit keys which has a total length of 192 bits. We can give more than one key that is two or three keys for encryption as well as for decryption such that the security will be stronger. It is times stronger than the normal DES algorithm, so that this algorithm can avoid the brute force attack. The main drawback of using 3DES algorithm is that the number of calculations is high reducing the speed to a greater extent. And the second drawback is that both DES and 3DES use same 64 block size to avoid security issues. “Advanced Encryption Standard” algorithms are used to avoid these limitations.
Advanced Encryption Standards:
Advanced Encryption Standards (AES) takes a block of size 128 bits as input and produces the output block of same size. AES supports different key sizes like 128, 192 and 256 bit keys. Each encryption key size will change the number of bits and also the complexity of cipher text.
The major limitation of AES is error propagation. The encryption operation and key generation both engage in number of non linear operations, so, for lengthy operations it is not suitable.

IV.VISUAL CRYPTOGRAPHY

Visual Cryptography is a special encryption technique to hide information in images in such a way that it can be decrypted by the human vision if the correct key image is used.
Specifically, visual cryptography allows effective and efficient secret sharing between a number of trusted parties. As with many cryptographic schemes, trust is the most difficult part. Visual cryptography provides a very powerful technique by which one secret can be distributed into two or more shares. When the shares are xeroxed onto transparencies and then superimposed exactly together, the original secret can be discovered without computer participation.
Supppose data D is devided into Nshhares then:

- D can be reconstructed from any k shares out of n
- Complete knowledge of k-1 share reveals no information about D
- K of n share is necessary to reveal secret data

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

1.VCS Algorithms

VCS Scheme normally involves two algorithms [4]:

- Algorithm for creating shares
- Algorithm for combining shares

One important functional requirement of any VCS system is size of shares which should be same as that of original image to prevent doubt for unauthorized user.

1.1 Algorithm for creating shares:

This algorithm divides secret image into n number of shares. The shares created by this algorithm will be in unreadable format such that it is impossible to reveal secret image. Single share cannot reveal the secret image. If these individual shares are transmitted separately through communication network, security is achieved.

1.2. Algorithm for combining shares:

This algorithm reveals the secret image by taking the number of shares as input. Some algorithm may take all shares as input and some other algorithm may take subset of shares as input. Decryption is done by merging shares which has taken as input.

DATA TRANSMISSION OVER NETWORK:

Wi‑Fi Protected Access (WPA and WPA2):

Wi‑Fi Protected Access encrypts information and makes sure that the network security key has not been modified. Wi‑Fi Protected Access also authenticates users to help ensure that only authorized people can access the network.

There are two types of WPA authentication: WPA and WPA2. WPA is designed to work with all wireless network adapters, but it might not work with older routers or access points. WPA2 is more secure than WPA, but it will not work with some older network adapters. WPA is designed to be used with an 802.1X authentication server, which distributes different keys to each user. This is referred to as WPA-Enterprise or WPA2-Enterprise. It can also be used in a pre-shared key (PSK) mode, where every user is given the same passphrase. This is referred to as WPA-Personal or WPA2-Personal.

VI.CONCLUSION AND FUTURE WORK

In the present world, the data transfers using internet is rapidly growing because it is so easier as well as faster to transfer the data to destination. So, many individuals and business people use to transfer business documents, important information using internet. Security is an important issue while transferring the data using internet because any unauthorized individual can hack the data and make it useless or obtain information un- intended to him.

The proposed approach in this project uses a new steganographic approach called image steganography. The application creates a stego image in which the personal data is embedded and is protected with a password which is highly secured.

The main intention of the project is to develop a steganographic application that provides good security. The proposed approach provides higher security and can protect the message from stego attacks. The image resolution doesn‟t change much and is negligible when we embed the message into the image and the image is protected with the personal password. So, it is not possible to damage the data by unauthorized personnel.

Using Least Significant Bit algorithm in this project for developing the application which is faster and reliable and compression ratio is moderate compared to other algorithms.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

The major limitation of the application is designed for bit map images (.bmp). It accepts only bit map images as a carrier file, and the compression depends on the document size as well as the carrier image size.

The future work on this project is to improve the compression ratio of the image to the text. This project can be extended to a level such that it can be used for the different types of image formats like .bmp, .jpeg, .tif etc., in the future. The security using Least Significant Bit Algorithm is good but we can improve the level to a certain extent by varying the carriers as well as using different keys for encryption and decryption.

VII.REFERENCES:

[1] Shyamalendu Kandar, Arnab Maiti, Variable Length Key based Visual Cryptography Scheme for Color Image using Random Number, International Journal of Computer Applications . Volume 19– No.4, April 2011.

[2] Ravindra Gupta, Akanksha Jain, Gajendra Singh, "Combine use of Steganography and Visual Cryptography for Secured Data hiding in Computer Forensics" , International Journal of Computer Science and Information Technologies, Vol. 3 (3) , 2012,4366 – 4370.

[3] Fridrich, J., Goljan, M. and Du,R, Reliable Detection of LSB Steganography in Colour and Grayscale Images, Proceedings of ACM Workshop volume 02, Manuscript Code: 11011on Multimedia and Security, Ottawa, October 5, 2001, pp.27-30.

[4] Talal Mousa Alkharobi, Aleem Khalid Alvi, New Algorithm for Halftone Image Visual Cryptography, IEEE 2004.

[5] Mrs.G.Prema and S.Natarajan, "Steganography using Genetic Algorithm along with Visual Cryptography for Wireless Network Application", IEEE 2012.

[6] R. Chandramouli, Nasir Menon, Analysis of LSB Based Image Steganography techniques, IEEE-2001.

[7] Arezoo Yadollahpour, Hossein Miar Naimi, Attack on LSB Steganography in Colour and Grayscale Images Using Autocorrelation Coefficients, European Journal of Scientific Research ISSN 1450- 216X Vol.31 No.2 (2009).

[8] Ghascmi E shanbchzadch J and ZahirAzami B, " A Steganographyc method based on Integer Wavelet Transform and Genetic Algorithm International Conference on Communications and Signal Processing (ICCSP) pp 42 45,2011.

[9] Dr.M.Umamaheswari Prof. S.Sivasubramanian S.Pandiarajan, Analysis of Different Steganographic Algorithms for Secured Data Hiding, IJCSNS International Journal of Computer Science and Network Security, VOL.10 No.8, August 2010

[10] Shyamalendu Kandar, Arnab Maiti, Variable Length Key based Visual Cryptography Scheme for Color Image using Random Number, International Journal of Computer Applications (0975 – 8887) Volume 19– No.4, April 2011

[11] Anupam Kumar Bairagi, ASCII based Even-Odd Cryptography with Gray code

and Image Steganography: A dimension in Data Security, ISSN 2078-5828 (Print), ISSN 2218-5224 (Online), Volume 01, Issue 02, Manuscript Code: 110112

Performance Evaluation of Statistical Approaches for Automatic Text-Independent Gender Identification System

R. Rajeswara Rao*
*Department of CSE, JNTUK Kakinada-Vizianagaram, AP, India

ABSTRACT

In this paper, robust feature for Automatic text-independent Gender Identification System has been explored. Through different experimental studies, it is demonstrated that the timing varying speech related information can be effectively captured using Hidden Markov Models (HMMs) than Gaussian Mixture Models (GMMs) . The study on the effect of feature vector size for good Gender Identification demonstrates that, feature vector size in the range of 18-22 can capture Gender related information effectively for a speech signal sampled at 16 kHz, it is established that the proposed Gender Identification system requires significantly less amount of data during both during training as well as in testing. The Gender Identification study using robust features for different states and different mixtures components, training and test duration has been exploited on TIMIT database.

Keywords - ***Gaussian Mixture Model (GMM),Ergodic Hidden Markov Models (EHMM) Gender, LPC, MFCC.***

I. INTRODUCTION

With the development of more and more identification systems to identity a Gender, there is a need for the development of a system which can provide identification task such as gender identification automatically without any human interface. Gender identification using voice of a person is comparatively easier than that from other approaches. There exist several algorithms for automatic gender identification but none of them has found to be 100% accurate. Gender Identification System can be represented like any other pattern recognition system as shown in Fig. 1. This task involves three phases, feature extraction phase, training phase and testing phase [1]. Training is the process of familiarizing the system with the voice characteristics of a speaker, whereas testing is the actual recognition task.

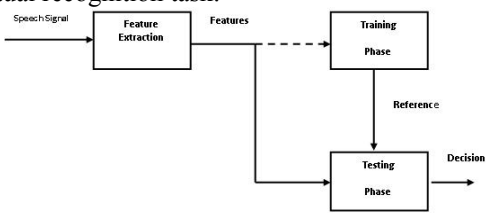


Fig. 1: A typical Block diagram representation of a Gender Identification task.

In Gender identification based on the voice of a speaker consists of detecting if a speech signal is uttered by a male or a female. Automatically detecting the gender of a speaker has several potential applications. In the context of Automatic Speech Recognition, gender dependent models are more accurate than gender independent ones [1] [2]. Hence, gender recognition is needed prior to the of speaker recognition. In the context of speaker recognition, gender detection can improve the performance by limiting the search space to speakers from the same gender. Also, in the context of content based multimedia indexing the speaker's gender is a cue used in the annotation. Therefore, automatic gender detection can be a tool in a content-based multimedia indexing system.

Much information can be inferred form a speech, such as sequences of words, gender, age, dialect, emotion, and even level of education, height or weight etc. Gender is an important characteristic of a speech. Automatically detecting the gender of a speaker has several potential applications such as (1) sorting telephone calls by gender (e.g. for gender sensitive surveys), (2) as part of an automatic speech recognition system to enhance speaker adaptation, and (3) as part of automatic speaker recognition systems. In the past, many methods of gender classification have been proposed. For parameters selections, some methods used gender dependent features such as pitch and formants [3] [5].

Speech is composite signal which has information about the message, gender, the speaker identity and the language [6][7]. It is difficult to isolate the speaker specific features alone from the signal. The speaker characteristics present in the signal can be attributed to the anatomical and the

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

behavioral aspects of the speech production mechanism. The representation of the behavioral characteristics is a difficult task, and usually requires large amount of data. Automatic speaker recognition systems rely mainly on features derived from the physiological characteristics of the speaker.

Speech is produced as sequence of sounds. Hence the state of vocal folds, shape and size of various articulators, change over time to reflect the sound being produced. To produce a particular sound the articulators have to be positioned in a particular way. When different speakers try to produce same sound, through their vocal tracts are positioned in a similar manner, the actual vocal tract shapers will be different due to differences in the anatomical structure of the vocal tract. System features represent the structure of vocal tract. The movements of vocal folds vary from one speaker to another. The manner and speed in which the vocal folds close also varies across speakers. Hence different voices are produced. Source features represent these variations in the vibrations of the vocal folds.

The theory of Linear Prediction (LP) is closely linked to modeling of the vocal tract system, and relies upon the fact that a particular speech sample may be predicted by a linear combination of previous samples. The number of previous samples used for prediction is known as the order of the prediction. The weights applied to each of the previous speech samples are known as Linear Prediction Coefficients (LPC). They are calculated so as to minimize the prediction error. As a byproduct of the LP analysis, reflection coefficients and log area coefficients are also obtained [8].

A study into the use of LPC for speaker recognition was carried out by Atal [9]. These coefficients are highly correlated, and the use of all prediction coefficients may not be necessary for speaker recognition task [10]. Sambur [11] used a method called orthogonal linear prediction. It is shown that only a small subset of the resulting orthogonal coefficients exhibits significant variation over the duration of an utterance. It is also shown that reflection coefficients are as good as the other feature sets. Naik et. al., [12] used principal spectral components derived from linear prediction coefficients for speaker verification task. Hence a detailed exploration to know the speaker-specific excitation information present in the residual of speech is needed and hence the motivation for the present work.

I. EXPLORING ROBUST FEATURES FOR GENDER IDENTIFICATION

Here, the GMM is used as front-end to extract features vectors from speech signal. For the Gender Identification ASR task, the basic requirement is to obtain the feature vectors form the speech signal. Recently, some attempts are made to explore the alternative representation of feature vectors based on GMM feature extraction.

For Speaker Recognition task, robust features are derived from the speech signal based on estimating a Gaussian mixture model. The underlying speaker discrimination information is represented by Gaussians. The estimated GMM parameters means, co-variance and component weight can be related to the formant locations, bandwidths and magnitudes.

For the proposed new feature vectors, from the speech signal of a speaker S_i, a 12 dimensional MFCC feature vectors are obtained with a window size of 20ms and window shift of 3 ms. These MFCC feature vectors are distributed into 'R' Gaussians mixtures as shown in Fig. 2.

Fig. 2: R Gaussians for Speaker S_i.

The feature vector X=(X1, X2,……, X12) is passed through a Gaussian G1 by calculating a Gaussian probability P1 using Gaussian probability density function. This P1 is first coefficient in the new feature vector. In the same way feature vector X is passed through R Gaussians by creating R feature vector coefficients namely P1,P2,….,PR, as shown in Fig. 3. These R coefficients create a new R dimensional feature vector. The newly created R dimensional feature vector is shown in the Fig. 4.

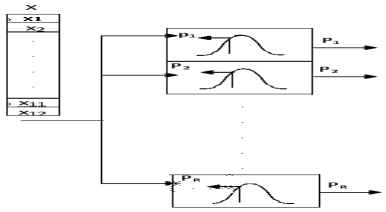


Fig. 3: Parameter estimation for new vector P.

When R=14, the optimal recognition performance has been achieved.

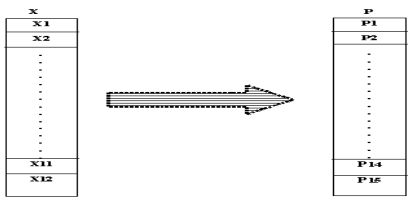


Fig. 4: Transforming from 12 dimensional MFCC feature vector to R dimensional feature vector.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Experiments are carried to find the dimension new feature vector for good speaker recognition performance. This is done by varying the number of Gaussians from 12 to 30, i.e. number of coefficients in the new feature vectors. When the numbers of coefficients are 20, the good identification performance is achieved [4].

II. Performance Evaluation of Statistical Approaches

A. Gaussian Mixture Model for Gender Identification

GMM is a classic parametric method best used to model gender identities due to the fact that Gaussian components have the capability of representing gender information effectively. Gaussian classifier has been successfully employed in several text-independent gender identification applications. As shown in Fig. 5 in a GMM model, the probability distribution of the observed data takes the form given by the following equation [13][14].

$$p(\bar{x} \mid \lambda) = \sum_{i=1}^{M} p_i b_i(\bar{x})$$

Where M is the number of component densities, $\bar{x}$ is a D dimensional observed data (random vector), $b_i(\bar{x})$ are the component densities and p_i are the mixture weights for i = 1, .., M.

$$b_i(\bar{x}) = \frac{1}{(2\pi)^{D/2} |\sum_i|^{1/2}} \exp\left\{-\frac{1}{2}(\bar{x}-\bar{\mu}_i)^T \sum_i^{-1} (\bar{x}-\bar{\mu}_i)\right\}$$

Each component density $b_i(\bar{x})$ denotes a D-dimensional normal distribution with mean vector $\bar{\mu}_i$ and covariance matrix $\sum_i$. The mixture weights satisfy the condition $\sum_{i=1}^{M} p_i = 1$ and therefore represent positive scalar values. These parameters can be collectively represented as $\lambda = \{p_i, \bar{\mu}_i, \Sigma_i\}$ for i = 1 … M. Each language in a language system can be represented by a GMM and is referred by the language respective model λ.

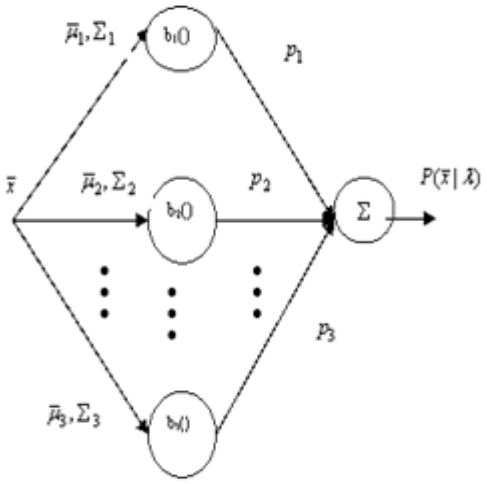


Fig. 5: Gaussian Mixture Model for Gender Identification

The parameters of a GMM model can be estimated using maximum likelihood (ML) [15] estimation. The main objective of the ML estimation is to derive the optimum model parameters that can maximize the likelihood of GMM. Unfortunately direct maximization using ML estimation is not possible and therefore a special case of ML estimation known as Expectation-Maximization (EM) [15] algorithm is used to extract the model parameters.

The GMM likelihood of a sequence of T training vectors $X = \{\bar{x}_1, ...\bar{x}_T\}$ can be given as [15].

The EM algorithm begins with an initial model λ and tends to estimate a new model $\bar{\lambda}$ such that

$p(X \mid \bar{\lambda}) \geq p(X \mid \lambda)$ [14]. As shown in Fig. 6, this is an iterative process where the new model is considered to be an

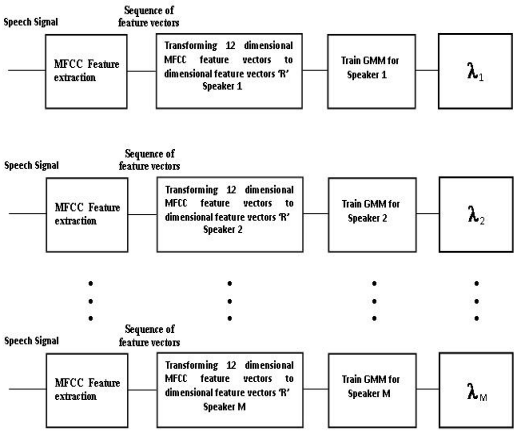


Fig. 6: Training GMM for Gender Identification Task

Initial model in the next iteration and the entire process is repeated until a certain convergence threshold is obtained

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

B.Continuous Ergodic Hidden Markov model for speaker recognition

The HMM is a doubly embedded stochastic process where the underlying stochastic process is not directly observable. HMMs have the capability of effectively modeling statistical variations in spectral features. In a variety of ways, HMMs can be used as probabilistic speaker models for both text-dependent and text-independent speaker recognition [17][18]. HMM not only models the underlying speech patterns but also the temporal sequencing among the sounds. This temporal modeling is advantageous for text-dependent speaker recognition system. Left Right HMM can model temporal sequence of patterns only, where as to capture the patterns of different type ergodic HMM is used [19]

As shown in the Fig. 4 in the training phase, one HMM for each speaker is obtained (i.e., parameters of model are estimated) using training feature vectors. The parameters of HMM are [MA, et.al, 2007] State-transition probability distribution: It is represented by $A = \lfloor a_{ij} \rfloor$

Where

$$a_{ij} = P(q_{t+1} = j \mid q_t = i) \quad 1 \le i, j \le N \tag{2}$$

defines the probability of transition from state i to j at time t.

For a three state left-right model the state transition matrix is given as

$$A = \{a_{ij}\} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{bmatrix} \tag{3}$$

The state transition matrix of three state ergodic model is given by

$$A = \{a_{ij}\} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{4}$$

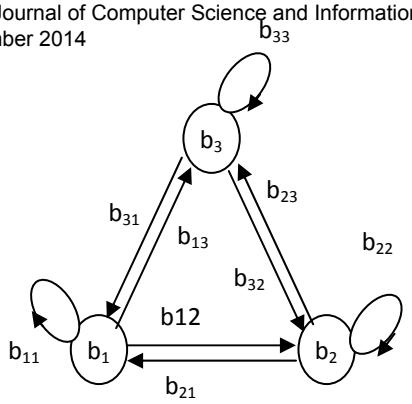


Fig. 6:Three-state ergodic HMM.

Observation symbol probability distribution: It is given by $B = \lfloor b_j(k) \rfloor$ in which

$$b_j(k) = P(O_t = V_k \mid q_t = j) \quad 1 \le k \le M \tag{5}$$

defines the symbol distribution in state $j = 1,2,3.....N$. The initial state distribution is given by $\pi = P(q_1 = i)$ where

$$\pi_i = P(q_1 = i) \quad 1 \le i \le N \tag{6}$$

ere, N is the total number of states, and q_t is the state at time t, M is the number of distinct observation symbols per state, and O_t is the observation symbol at time t. In testing phase, $P(O/\lambda)$ for each model is calculated, where $O = (O_1 O_2 O_3 ....O_T)$ Here the goal is to find out the probability for a given model to which the test utterance belongs to. The speaker whose model gives the highest score is declared as the identified speaker. GMM corresponds to a single-state continuous ergodic HMM.

The model parameters can be collectively represented as $\lambda = (A_i, B_i, \pi_i)$ for $i = 1.....M$. Each speaker in a speaker identification system can be represented by a HMM and is referred to by the speaker's respective models λ.

In the testing phase, p (O/λ) for each model is calculated [21]. where O=(o1o2o3…OT) is the sequence of the test feature vectors. The goal is to find the probability, given the model, that the test utterance belongs to that particular model. The speaker model that gives the highest score is declared as the ident

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

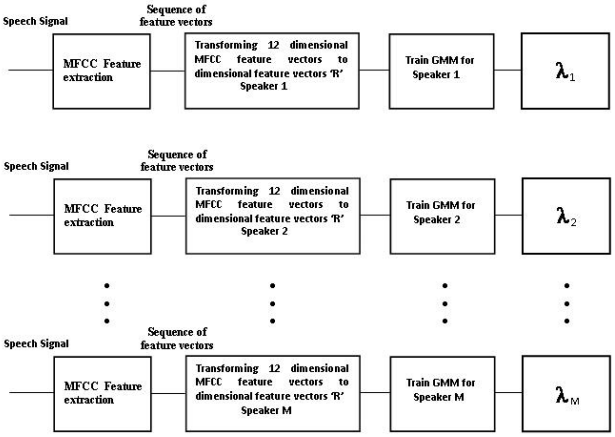


Fig. 6: Training HMM for Gender Recognition Task

estimation is not possible and therefore a special case of ML estimation known as Expectation-Maximization (EM) [K. N. Stevens, 1999] algorithm is used to extract the model parameters.

The GMM likelihood of a sequence of T training vectors $X = \{\overline{x}_1, \ldots \overline{x}_T\}$ can be given as [16]

$$p(X \mid \lambda) = \prod_{t=1}^{T} p(\overline{x}_t \mid \lambda)$$

The EM algorithm begins with an initial model λ and tends to estimate a new model $\overline{\lambda}$ such that $p(X \mid \overline{\lambda}) \geq p(X \mid \lambda)$ [21]. This is an iterative process where the new model is considered to be an initial model in the next iteration and the entire process is repeated until a certain convergence threshold is obtained

III. EXPERIMENTAL EVALUATION

A. Database used for the study

Gender identification is the task of identifying whether the speaker is male or female. In this paper we consider identification task for TIMIT Speaker database [16].

The TIMIT corpus of read speech has been designed to provide speaker data for the acquisition of acoustic-phonetic knowledge and for the development and evaluation of automatic speaker recognition systems. TIMIT contains a total of 6300 sentences, 10 sentences spoken by each of 630 speakers from 8 major dialect regions of the United States. We consider 100 male speakers and 100 female out of 630 speakers for gender recognition. Maximum of 30 sec. of speech data is used for training and minimum of 1 sec. of data for testing. In all the cases the speech signal was sampled at 16 kHz sampling frequency. Throughout this study, closed set identification experiments are done to demonstrate the feasibility of capturing the Gender - discrimination information from the speech signal. Requirement of significantly less amount data for Gender-discrimination information and Gaussian mixture models is also demonstrated.

B. Experimental Setup

The system has been implemented in Matlab7 on Windows XP platform. We have trained the model GMM using Gaussian Components as 2, 4, 8, and 16 for training speech duration of 10, 20 and 30 sec. Testing is performed using different test speech durations such as 1 sec., 2 sec., and 3 sec..

II. Performance Evaluation

The system has been implemented in Matlab7 on windows XP platform. The result of the study has been presented in Table 1. We have used Vector order of 18 for all experiments. We have trained the model using Gaussian mixture components as 4, 8, 16, 32 and 64 for training speech lengths as 20 sec.,. Testing is performed using different test speech lengths such as 1 sec, 3 sec, and 5 sec.. Here, recognition rate is defined as the ratio of the number of genders identified to the total number of genders tested. As shown in Table. 1 the identification rate for testing length for 5 sec. outperformed, where as for testing length of 3 sec. is also on par with 5 sec. testing length. Table. 1, shows identification rate increases when different number of mixture components 4, 8, 16, 32 and 64 with different test speech lengths 1 sec., 3 sec., and 5 sec..

The percentage (%) recognition of Gaussian Components such as 4, 8, 16, 32 and 64 seems to be uniformly increasing. The minimum number of Gaussian components to achieve good recognition performance seems to be 32 and thereafter the recognition performance is minimal. The recognition performance of the HMM drastically increases for the test speech duration of 1 sec. to 3 sec.. Increasing the test speech duration from 3 sec. to 5 sec. improves the recognition performance with small improvement.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Table 1: Gender Recognition Performance for 20 Sec. Training speech duration

No. of States	No. of Mixture Components	Speaker Recognition (%) Test Duration (in sec.)		
		1 Sec.	3 Sec.	5 Sec.
2	4	74	88	94
	8	82	95	98
	16	84	96	98
	32	86	97	99.5
	64	84	94	97
3	4	96	98	98.5
	8	98	98.5	100
	16	98.5	100	100
	32	99	99.5	99
	64	97	98	98.5
4	4	95	96	98
	8	94	96.5	98
	16	96	98.5	99
	32	97	98	99
	64	95	97	99

Table 2: Table 1: Gender Identification Performance

Training speech duration (sec)	No. of mixture components	Recognition rate (%) Testing speech length		
		1 sec	3 sec	5 sec
10	2	91	92	93
	4	93	93.5	94
	8	93.5	93	94.5
	16	93.5	93	94
20	2	93	94	94
	4	93.5	94.5	95
	8	94	95	95.5
	16	94	96	97
30	2	94	96	96.5
	4	94.5	96.5	97
	8	95	97	98.5
	16	96	98	99

IV. CONCLUSION

In this work we have demonstrated the importance of coefficient order for speaker recognition task. gender discrimination information is effectively captured for coefficient order 18 using a HMM than GMM. The recognition performance depends on the training speech length selected for training to capture the gender-discrimination information. Larger the training length, the better is the performance, although smaller number reduces computational complexity.

The objective in this paper was mainly to demonstrate the significance of the gender-discrimination information present in the speech using stastical approaches. We have not made any attempt to optimize the parameters of the model used for feature extraction, and also the decision making stage. Therefore the performance of speaker recognition may be improved by optimizing the various design parameters.

References

[1] Alex Acero and Xuedong Huang, "*Speaker and Gender Normalization for Continuous-Density Hidden Markov Models"*, in Proc. of the Int. Conf. on Acoustics, Speech, and Signal , IEEE, May 1996.

[2] C. Neti and Salim Roukos., "*Phone-specific gender-dependent models for continuous speech recognition"*, Automatic Speech Recognition and Understanding Workshop (ASRU97), Santa Barbara, CA, 1997.

[3] R. Vergin, A. Farhat and D.O'Shaughnessy, "*Robust gender-dependent acoustic-phonetic modeling in continuous speech recognition based on a new automatic male/female classification*", Proc. Of IEEE Int. Conf. on Spoken Language (ICSLP), pp. 1081, Oct. 1996.

[4] A. Nagesh, V. Kamakshi Prasad, " New Feature Vectors for Automatic Text-Independet Language Identification" IJCSC, Vol. 2, Issue 2 pp553-558, 2011.

[5] S. Slomka and S. Sridharan, "*Automatic gender identification optimized for language independence*", Proc. Of IEEE TENCON'97, pp. 145-148,Dec. 1997.

[6] O'Shaughnessy D., 1987.,"*Speech Communication: Human and Machine"*, Addison-Wesley, New York.

[7] Rabiner L.R., Juang B.H., 1993. "*Fundamentals of Speech Recognition",* Prentice-Hall, Englewood Cliffs, NJ.

[8] Makhoul, J., 1975. "*Linear prediction: a tutorial review",* Proc. IEEE 63, 561–580.

[9] B.S. Atal, "*Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification*", J. Acoust. Soc. Ameri., *vol. 55*, pp.1304-1312, Jun. 1974.K. Elissa, "Title of paper if known," unpublished.

[10] A.E. Rosenberg and M. Sambur, "*New techniques for automatic speaker verification*", *vol. 23, no.2*, pp.169-175, 1975.

[11] M. R. Sambur, "*Speaker recognition using orthogonal linear prediction*", IEEE Trans. Acoust. Speech, Signal Processing, *vol. 24*, pp.283-289, Aug. 1976

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

[12] J. Naik and G. R. Doddington, " *High performance speaker verification using principal spectral components*", in proc. IEEE Int. Conf. Acoust. Speech, Singal Processing, pp. 881-884, 1986.

[13] Gish H., Krasner M., Russell W., and Wolf J., "*Methods and experiments for text-independent speaker recognition over telephone channels*", Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), *vol. 11*, pp. 865-868, Apr. 1986.

[14] Reynolds D. A., and Rose R. C., " *Robust Text-Independent Speaker Identification using Gaussian Mixture Models*'', IEEE-Transactions on Speech and Audio Processing, *vol. 3, no. 1*, pp. 72-83,1995.

[15] A. P. Dempster, N. M. Laird, and D. B. Rubin, "*Maximum likelihood from incomplete data via the EM algorithm*", J. Royal Statist. Soc. Ser. B. (methodological), *vol. 39*, pp. 1-38, 1977

[16] K.N. Stevens, *Acoustic Phonetics*. Cambridge, England: The MIT Press, 1999

[17] M. Forsyth, "Discriminating observation probability (dop) HMM for speaker verification", Speech Comm., vol. 17, pp. 117-129, 1995.

[18] R, Rajeshwara Rao, "Automatic Text - Independent Speaker Recognition using source features", Ph.D. thesis., Jan-2010.

[19] L. R. Rabiner and B. H. Juang, Fundamentals of Speech Recognition. Prentice-Hall, 1993.

[20] MA, B., Li, H., and Tong, R. Spoken language recognition with ensemble classifier. IEEE Trans. Audio, Speech and Language Processing 15, 7 (September 2007), 2053-2062.)

[21] A. P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm", J. Royal Statist. Soc. Ser. B. (methodologies), vol. 39, pp. 1-38, 1977.

[22] K.N. Stevens, Acoustic Phonetics. Cambridge, England: The MIT Press, 1999.

`

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

Book Related Information Retrieval Using Ontology Based Semantic Search and WordPress

P B Lahoti
Computer Science & Engineering
Government College of Engineering
Aurangabad, India.

V A Chakkarwar
Computer Science & Engineering
Government College of Engineering
Aurangabad, India.

***Abstract*-Nowadays most of the contents are stored and shared on the web, so it is difficult to an intelligent user to find the exact content when it comes to finding and properly managing information in massive volumes. This difficulties are occurred due to Traditional keyword search engine model. Aiming to solve the limitations of keyword-based search engine Semantic Web (SW) has been introduced. Main idea of this research paper is to explore the current state of the semantic information retrieval with major focus on ontology based search. The research paper includes introductory knowledge on the Semantic Web and its layer cake, proposed method which make use of WordPress which is a free and open source blogging tool and a content management system (CMS) based on PHP and MySQL**.

Keywords-Semantic retrieval, Ontology based Information search, Book search, WORDPRESS, Ontology based Data Integration (ONDINE), MIEL++.

I. INTRODUCTION

The amount of information available in World Wide Web (WWW) is very large and still growing, which makes retrieval of information from WWW a tedious task. There are many search engines developed to address this problem, but most of them adopt the traditional keyword based search. Keyword based search method uses the user query to retrieve set of relevant documents from the indexed document those fit the terms given by the user. Semantic Web is an extension of current web in which information provides well-defined meaning that enables system and people for better understanding and can enable to work effectively by understanding information from different sources [1]. The introduction of semantic web is a great leap from the existing Web 2.0 in which the user not only interacts with the web, but also has the capability to generate more meaningful information. The complete information is represented with the help of Ontology. Ontology allows knowledge to be represented as a set of concepts, properties and the relations between them.

In information retrieval, the users don't search with the exact terms represented in the documents in most of the cases. Hence, relevant documents are not fetched by the keyword-based information retrieval but the semantic web makes the information retrieval more users driven than that of keyword driven. Hence it helps to retrieve more relevant documents.

The capabilities of current software to interpret web content and extract useful information are very limited. An alternative approach is to represent web content in a form that is easily processed by machines. This plan to revolutionize the web is semantic web initiative.

A. *Semantic Web*

Semantic Web has become a current challenge in World Wide Web (WWW), where it will lead to a new type of sharing data on the net openly. It has been described in rather different ways: as a utopic vision, as a web of data, or merely as a natural paradigm shift in our daily use of the Web. Most of all, the Semantic Web has inspired and engaged many people to create innovative semantic technologies and applications.

B. *Ontology*

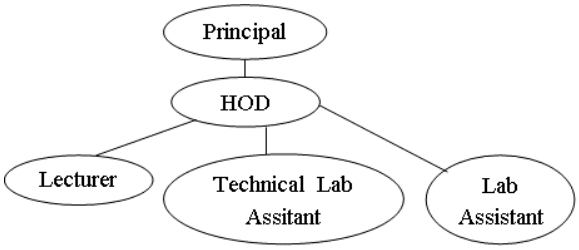


Fig.1. Ontology for faculty of college

It is a specification of all the relevant concepts and their relationships within a given domain, typically in a hierarchical data structure. A common set of terms that describes and represents a domain is defined as ontology. It can enhance the functioning of web by improving the accuracy of web searches.

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

C. Ontology

Ontology is a "formal, explicit specification of a shared conceptualization of a domain of interest". Thus, Ontology is the attempt to express an exhaustive conceptual scheme within a given domain, typically a hierarchical data structure containing all the relevant entities, their relations and the rules within that domain [8].

An Ontology is a 5-tuples O = (C, P, R, I, A), where:

C represents classes or domain concepts and can be arranged in inheritance hierarchies. They should give the specific definition of concepts both in syntax and semantics level.

- P is a set of concept properties.
- R is a set of binary semantic relations defined between concepts {one-to-one, one-to-many, many-to-many} is the set of relation type.

A set of basic relations is defined as:

R= {synonym-of, kind-of, part-of, instance-of}, which have the following interpretations

- Part-of relation depicts relation of part and integrity between two concepts.
- Kind-of relation is represented by characteristics of inheritance relationship of two concepts.
- Instance-of relation describes inclusion relationship between a concept and its subordinate instance.
- Synonym-of relation depicts equivalence relation between concepts.
- A is a set of axioms. An axiom is a real fact or reasoning rules.

II. RELATED WORK

Semantic Search tends to improve retrieval effectiveness. Guha et al [2] designed an application called Semantic Search to improve traditional web searching. Based on the scope of semantic search, it has been applied in different environments. Finin et al [3] discussed about applying semantic search over web so that it improves search effectiveness of information retrieval systems. In Semantic Web area, semantic search system provides search mechanisms over a single KB which is different from standard Information Retrieval (IR) model that provides document searching. Hence, there is more emphasis on developing new techniques that captures user queries and converts them into formal query representation.

Fernandez et al [4] designed a retrieval system which follows ontology based semantic search approach. The overall retrieval process of the system consists of following steps. The system takes natural language query as input and it is converted into semantic entities by query processing module which has been replaced by cross-ontology question answering system, PowerAqua. The second step is to retrieve and rank the documents related to users query. For this, documents that are annotated are indexed for retrieval purpose using indexing module which consists of annotation algorithm. The final output of the system is a complementary list of semantically ranked relevant documents and a set of ontology elements that answer user question.

Castells et al [5] designed a retrieval system that exploits ontology based KBs to improve search over large document repositories. Semantic search is combined with traditional keyword based retrieval which tolerates sparseness of KB. The overall retrieval process consists of following steps.

This system takes as input RDF Data Query Language (RDQL) query and this is executed against the KB. The output of this step is list of instance tuples that satisfy the query. For this execution, ontology processing library, Jena Toolkit is used. Document Annotation is done using semiautomatic technique. These annotations are given weights based on TFIDF algorithm. The documents that are annotated with the instances returned in previous step are presented to the user. Giunchiglia et al [6] presented an approach called concept search which is search based on computation of semantic relation between concepts. It reuses retrieval model and data structures of syntactic search but the only difference is that words are replaced with concepts and syntactic matching of words is extended to semantic matching of concepts.

The semantic resource used for most of the query answer-ing systems is ontology. One such system called PowerAqua, designed by Lopez et al [7] takes as input natural language query and returns answers retrieved from ontologies found anywhere on semantic web.

Latifur Khan, Dennis McLeod, Eduard Hovy [9] worked on the key problem in achieving efficient and user friendly retrieval is the development of a search mechanism to guarantee delivery of minimal irrelevant information (high precision) while insuring relevant information is not overlooked (high recall). To achieve this, they proposed a potentially powerful and novel approach for the retrieval of audio information. In their research they explained the development of an ontology-based model for the generation of metadata for audio, and the selection of audio information in a user customized manner. Also conclude how the ontology they proposed can be used to generate information selection requests in database queries. Vaclav Snasel, Pavel Moravec, Jaroslav Pokorny [10 presented a basic method of mapping LSI concepts on given ontology (WordNet), used both for retrieval recall improvement and dimension reduction. They offered experimental results for this method on a subset of TREC collection, consisting of Los Angeles Times articles. In their research they had shown, that mapping terms on WordNet hypernyms improves recall, bringing more relevant documents. The LSI filtration enhances recall even more, producing smaller index, too. The question is, whether use expensive method as LSI just for the term filtration. The third approach – using LSI on generated hypernym-by-document matrix has yet to be tested.

Sofia Stamou [11] had discussed keyword-based searching does not always result to the retrieval of qualitative data, basically due to the variety in the vocabulary used to convey alike information. In this paper, introduce a concept-based retrieval model, which tackles vocabulary mismatches through the use of domain-dependent ontologies. In particular,

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

our model explores the information encoded in domain ontologies for indexing documents according to their semantics rather than wordforms. To demonstrate the potential of proposed model built an experimental prototype which employs the topical ontologies for indexing Web documents in terms of their semantics. Zeng Dan [12] worked on Semantic Information Retrieval Based on Ontology to resolve the problem of the accuracy on traditional information retrieval, which brings ontology-based semantic information retrieval. The anthor wlilized the method of establishing the domain semantic model with ontology technology, the membership of concept added to the process of semantic modeling, and to provide semantic annotation to facilitate computer calculation processing. Qin Zhana Xia Zhang, Deren Li [13], proposed a approach to overcome the problems of semantic heterogeneity, the explication of knowledge by means of ontology, which can be used for the identification and association of semantically corresponding concepts because ontology can explicitly and formally represent concepts and relationships between concepts and can support semantic reasoning according to axioms in it. Ontology has been developed in the context of Artificial Intelligent (AI) to facilitate knowledge sharing and reuse. In this paper, an ontology-based semantic description model is put forward to explicitly represent geographic information semantics in abstract level and concrete level by introducing Ontologies.

Axel Reymonet, Jerome Thomas, Nathalie Aussenac-Gilles [14], presented a semantic search engine designed to handle within two separate tools both aspects of semantic IR: semantic indexing and semantic search. search engine only exploits knowledge explicitly mentioned in each request/document, the ability to express causal information in OWL could be taken into account in order to bring closer two symptoms apparently different but which share one (or more) fault(s) as potential origin for a given breakdown. Gaihua Fu, Christopher B. Jones and Alia I. Abdelmoty [15], the query expansion techniques presented in this paper are based on both a domain and a geographical ontology. Different from term-based query expansion techniques, the proposed techniques expand a query by trying to derive its geographical query footprint, and it is specially designed to resolve a spatial query. Various factors, such as types of spatial terms as encoded in the geographical ontology, types of non-spatial terms as encoded in the domain ontology, the semantics of the spatial relationships, their context of use, and satisfiability of initial search result, are taken into account to support expansion of a spatial query. The proposed techniques support the intelligent, flexible treatment of a spatial query when a fuzzy spatial relationship is involved. Some experiments have been carried out to evaluate the performance of the proposed techniques using sample realistic ontologies. Jan Paralic, Ivan Kostial [16], in the proposed model, a new, ontology-based approach to information retrieval (IR) is presented. The system is based on a domain knowledge representation schema in form of ontology. New resources registered within the system are linked to concepts from this ontology. In such a way resources may be retrieved based on the associations and not only based on partial or exact term matching as the use of vector model presumes. The ontology-based retrieval mechanism has been compared with traditional full text search based on vector IR model as well as with the Latent Semantic Indexing method.

Stuart Aitken and Sandy Reid in this paper [17], evaluated the use of an explicit domain ontology in an information retrieval tool. The evaluation compares the performance of ontology-enhanced retrieval with keyword retrieval for a fixed set of queries across several data sets. The robustness of the IR approach is assessed by comparing the performance of the tool on the original data set with that on previously unseen data. The empirical evaluation of ontology-based retrieval in CB-IR has broadly confirmed the hypotheses about relative and absolute performance of the system and about the adequacy and robustness of the ontology. Asunción Gómez-Pérez, Fernando Ortiz-Rodríguez, Boris Villazón-Terrazas [18], worked on “Ontology-Based Legal Information Retrieval to Improvethe Information Access in e-Government”. In this paper, approach to an ontology-based legal IR, which aims to retrieve government documents in a timely an accurate way.

III. PROPOSED WORK

A framework for book ontology based Information Retrieval model that is expected to improve retrieval effectiveness has been proposed. This framework is depicted in Fig. 2.

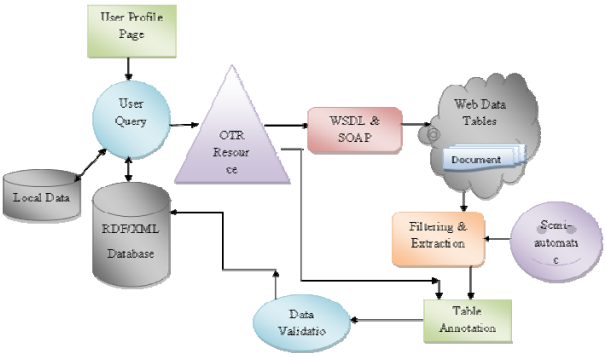


Fig.2. Semantic Retrieval System

The systems architect establishes the basic structure of the system, defining the essential core design features and elements that provide the framework. The systems architect provides the architects view of the users' vision. Above diagram shows that the user profile page and user query will be converted into search based on ontology & terminology based queries. Then, the OTR based data query will give to WSDL & SOAP process to retrieve the data from web documents. Later, the data which will be available in web data tables will be filtered & extracted by using of semi-automatic process and thereafter the data will be annotated based on the OTR based phase will be done & later it will validate the data to give the integrated output.

Here we make use of WordPress which is a free and open source blogging tool and a content management system (CMS) based on PHP and MySQL. Features include plugin architecture

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

and a template system. For Querying we make use of MIEL++ query which is asked by the end user into the XML/RDF data warehouse which contains fuzzy RDF graphs generated by our annotation method to annotate XML data tables, the query processing has to deal with fuzzy values.

Following are the modules used in the proposed system:

A. User Login & User Query Module

In this module, we are going to design web application to main originalities of our new flexible querying subsystem are:

- to retrieve not only exact answers compared with the selection criteria but also semantically close answers;
- to compare the selection criteria expressed as fuzzy sets representing preferences with the fuzzy annotations of data tables.

Querying subsystem allows the end-user to express preferences in his/her query and to retrieve the nearest data stored in the two kinds of data sources corresponding to his/her selection criteria.

B. Ontology & WordPress

Ontology is a "formal, explicit specification of a shared conceptualization of a domain of interest". Thus, Ontology is the attempt to express an exhaustive conceptual scheme within a given domain, typically a hierarchical data structure containing all the relevant entities, their relations and the rules within that domain. We build domain ontology of books, and then we present the semantic retrieval system of books information using WordPress software.

C. Filtering & Table Extraction

Recent propositions in the Semantic Web community propose to extract, filter, annotate and query Web data tables, but they have not been designed with the same objectives as ours. Table Seer for instance allows a set of predefined metadata to be extracted from Web data tables, but it does not compare the schema of the Web data tables with preexisting schemas defined in ontology. We can also cite Web Tables which proposes a system to identify relational tables in a huge amount of tables included in HTML documents and to index them, this in order to query and rank them.

D. Table Annotation With Ontology Based

Our method to identify relations depends on the identification of the symbolic concepts and quantities, which can be considered as a weakness. For this reason, our experimentation to automatically annotate the data tables with the relations of the considered OTR was applied without validating the intermediate steps.

E. Validation & Storing into RDF/XML Database

In this module, when a query is asked by the end user into the XML/RDF data warehouse which contains fuzzy RDF graphs generated by our annotation method to annotate XML data tables, the query processing has to deal with fuzzy values. More precisely, it has

- To take into account the certainty score associated with the relations represented in the data tables and
- To compare a fuzzy set expressing querying preferences to a fuzzy set, generated by our annotation method, having a semantic of similarity or imprecision.

IV. PAREMETERS FOR EVALUATION

A. Precision

Precision is one the most commonly used metrics in the IR world. It basically measures how precisely the system picks the related documents among all documents. More specifically, it is the proportion of the related documents in the retrieved documents (true positives) to the total number of retrieved documents (Eq. 2.1). Precision, on its own, does not give much information about the actual performance of the system, since it does not consider whether or not all the related documents are retrieved.

$$\text{Precision} = \frac{\text{true positive}}{\text{true positive + false positive}} \quad (1.1)$$

B. Recall

Recall is another widely used IR metric. It is the proportion of the retrieved related documents to the total number of related documents that should have been retrieved. Similar to precision, it is not much meaningful on its own, because it does not takes into account the unrelated documents retrieved (Eq. 2.2).

$$\text{Recall} = \frac{\text{true positive}}{\text{true positive + false negative}} \quad (1.2)$$

V. CONCLUSION

With the development of internet and the huge amount of data related to the book is increased today, so retrieval of relevant books is a challenging task. System is able to overcome the limitations of web 2.0 by representing the knowledge in ontology. Ontology represents the knowledge in terms of classes and subclasses. Knowledge represented in ontology can be interpreted by machine. Machine can add more data and relations on behalf of users. Due to increase in use of semantic web, number of ontology has been increased which created the problem for ontology storage. Problem was there for scalable storage of ontology data. System uses the different semantic matching operators for ontology search. Querying the semantic data is simplified because of relational databases. This paper has studied semantic retrieval of books information based on ontology and SPARQL. Then we have proposed books ontology model and information retrieval system which will perform intelligent information retrieval through semantic relationship

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

between books concept and this system also gives a mechanism to retrieve synonym words which is the major issue in WWW. In future, the same concept can be easily adapted for another domain by doing the certain changes in Information extraction, ontology design and database design. The concept can be used to build ontology for multilingual domain which collects the data from different language repository.

VI. REFERENCES

[1] Berners-Lee, J. Hendler, and O. Lassila, *The Semantic Web*, Scientific American, vol. 284-5, 2001, p. 34-43.

[2] Guha R.V., McCool R. and Miller E. (2003) Semantic search, Proceedings of the 12th International World Wide Web Conference, pp.700-709.

[3] Finin T., Mayfield J., Fink C., Joshi A. and Cost R.S. (2005) Information retrieval and the Semantic Web, Proceedings of the 38th Annual Hawaii International Conference on System Sciences, Big Island, HI, USA, pp.414. Belmont, CA: Wadsworth, 1993, pp. 123135.

[4] Fernandez M., Cantador I., Lopez V., Vallet D., Castells P. and Motta E. (2011) Semantically enhanced information retrieval:An ontology based approach ,Journal of Web Semantics,vol.9, pp.434-452.

[5] Castells P., Fernandez M. and Vallet D. (2007) An adaptation of vector space model for ontology-based information retrieval, IEEE Transactions on Knowledge and Data Engineering, vol.19, Issue 2, pp.261-272.

[6] Giunchiglia F., Kharkevich U. and Zaihrayeu I. (2009) Concept search, Proceedings of the 6th European Semantic Web Conference Heraklion, Greece, pp.429444.

[7] Lopez V., Sabou M. and Motta E. (2009) Crossontology question answering on semantic web-an initial evaluation, Proceedings of the knowledge Capture Conference, California, CA, USA, pp.17-24.

[8] M. Karyda, T. Balopoulos, et. Al. “An ontology for secure e-government applications”, Proceedings of the First International Conference on Availability, Reliability and Security (ARES’06) 0-7695-2567-9/06, 2006 IEEE.

[9] Latifur Khan, Dennis McLeod, Eduard Hovy, “Retrieval Effectiveness of an Ontology-Based Model for Information Selection”.

[10] Vaclav Snasel, Pavel Moravec, Jaroslav Pokorny, “WordNet Ontology Based Model for Web Retrieval”.

[11] Sofia Stamou, “Retrieval Effectiveness of An Ontology-Based Model For Conceptual Indexing”.

[12] Zeng Dan, “Research on Semantic Information Retrieval Based on Ontology”, Proceedings of the 7th International Conference on Innovation & Management, Page No. 1582-1586.

[13] Qin Zhana Xia Zhang, Deren Li, “Ontology-Based Semantic Description Model For Discovery And Retrieval Of Geo-Spatial Information”, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. XXXVII. Part B4. Beijing 2008, Page 141-146.

[14] Axel Reymonet, Jerome Thomas, Nathalie Aussenac-Gilles, “Ontology Based Information Retrieval:an application to automotive diagnosis”.

[15] Gaihua Fu, Christopher B. Jones and Alia I. Abdelmoty, “Ontology-based Spatial Query Expansion in Information Retrieval”.

[16] Jan Paralic, Ivan Kostial, “Ontology-based Information Retrieval”, EC funded project IST-1999-20364, Webocracy (Web Technologies Supporting Direct Participation in Democratic Processes).

[17] Stuart Aitken and Sandy Reid, “Evaluation of an Ontology-BasedInformation Retrieval Tool”.

[18] Asunción Gómez-Pérez, Fernando Ortiz-Rodríguez, Boris Villazón-Terrazas, “Ontology-Based Legal Information Retrieval to Improve the Information Access in e-Government”, WWW 2006, May 23–26, 2006, Edinburgh, Scotland. ACM 1-59593-323-9/06/0005

http://sites.google.com/site/ijcsis/
ISSN 1947-5500

IJCSIS AUTHORS' & REVIEWERS' LIST

Assist Prof (Dr.) M. Emre Celebi, Louisiana State University in Shreveport, USA
Dr. Lam Hong Lee, Universiti Tunku Abdul Rahman, Malaysia
Dr. Shimon K. Modi, Director of Research BSPA Labs, Purdue University, USA
Dr. Jianguo Ding, Norwegian University of Science and Technology (NTNU), Norway
Assoc. Prof. N. Jaisankar, VIT University, Vellore,Tamilnadu, India
Dr. Amogh Kavimandan, The Mathworks Inc., USA
Dr. Ramasamy Mariappan, Vinayaka Missions University, India
Dr. Yong Li, School of Electronic and Information Engineering, Beijing Jiaotong University, P.R. China
Assist. Prof. Sugam Sharma, NIET, India / Iowa State University, USA
Dr. Jorge A. Ruiz-Vanoye, Universidad Autónoma del Estado de Morelos, Mexico
Dr. Neeraj Kumar, SMVD University, Katra (J&K), India
Dr Genge Bela, "Petru Maior" University of Targu Mures, Romania
Dr. Junjie Peng, Shanghai University, P. R. China
Dr. Ilhem LENGLIZ, HANA Group - CRISTAL Laboratory, Tunisia
Prof. Dr. Durgesh Kumar Mishra, Acropolis Institute of Technology and Research, Indore, MP, India
Dr. Jorge L. Hernández-Ardieta, University Carlos III of Madrid, Spain
Prof. Dr.C.Suresh Gnana Dhas, Anna University, India
Dr Li Fang, Nanyang Technological University, Singapore
Prof. Pijush Biswas, RCC Institute of Information Technology, India
Dr. Siddhivinayak Kulkarni, University of Ballarat, Ballarat, Victoria, Australia
Dr. A. Arul Lawrence, Royal College of Engineering & Technology, India
Dr. Wongyos Keardsri, Chulalongkorn University, Bangkok, Thailand
Dr. Somesh Kumar Dewangan, CSVTU Bhilai (C.G.)/ Dimat Raipur, India
Dr. Hayder N. Jasem, University Putra Malaysia, Malaysia
Dr. A.V.Senthil Kumar, C. M. S. College of Science and Commerce, India
Dr. R. S. Karthik, C. M. S. College of Science and Commerce, India
Dr. P. Vasant, University Technology Petronas, Malaysia
Dr. Wong Kok Seng, Soongsil University, Seoul, South Korea
Dr. Praveen Ranjan Srivastava, BITS PILANI, India
Dr. Kong Sang Kelvin, Leong, The Hong Kong Polytechnic University, Hong Kong
Dr. Mohd Nazri Ismail, Universiti Kuala Lumpur, Malaysia
Dr. Rami J. Matarneh, Al-isra Private University, Amman, Jordan
Dr Ojesanmi Olusegun Ayodeji, Ajayi Crowther University, Oyo, Nigeria
Dr. Riktesh Srivastava, Skyline University, UAE
Dr. Oras F. Baker, UCSI University - Kuala Lumpur, Malaysia
Dr. Ahmed S. Ghiduk, Faculty of Science, Beni-Suef University, Egypt
and Department of Computer science, Taif University, Saudi Arabia
Dr. Tirthankar Gayen, IIT Kharagpur, India
Dr. Huei-Ru Tseng, National Chiao Tung University, Taiwan

CALL FOR PAPERS
International Journal of Computer Science and Information Security

IJCSIS 2015
ISSN: 1947-5500
http://sites.google.com/site/ijcsis/

International Journal Computer Science and Information Security, IJCSIS, is the premier scholarly venue in the areas of computer science and security issues. IJCSIS 2011 will provide a high profile, leading edge platform for researchers and engineers alike to publish state-of-the-art research in the respective fields of information technology and communication security. The journal will feature a diverse mixture of publication articles including core and applied computer science related topics.

Authors are solicited to contribute to the special issue by submitting articles that illustrate research results, projects, surveying works and industrial experiences that describe significant advances in the following areas, but are not limited to. Submissions may span a broad range of topics, e.g.:

Track A: Security

Access control, Anonymity, Audit and audit reduction & Authentication and authorization, Applied cryptography, Cryptanalysis, Digital Signatures, Biometric security, Boundary control devices, Certification and accreditation, Cross-layer design for security, Security & Network Management, Data and system integrity, Database security, Defensive information warfare, Denial of service protection, Intrusion Detection, Anti-malware, Distributed systems security, Electronic commerce, E-mail security, Spam, Phishing, E-mail fraud, Virus, worms, Trojan Protection, Grid security, Information hiding and watermarking & Information survivability, Insider threat protection, Integrity
Intellectual property protection, Internet/Intranet Security, Key management and key recovery, Language-based security, Mobile and wireless security, Mobile, Ad Hoc and Sensor Network Security, Monitoring and surveillance, Multimedia security ,Operating system security, Peer-to-peer security, Performance Evaluations of Protocols & Security Application, Privacy and data protection, Product evaluation criteria and compliance, Risk evaluation and security certification, Risk/vulnerability assessment, Security & Network Management, Security Models & protocols, Security threats & countermeasures (DDoS, MiM, Session Hijacking, Replay attack etc,), Trusted computing, Ubiquitous Computing Security, Virtualization security, VoIP security, Web 2.0 security, Submission Procedures, Active Defense Systems, Adaptive Defense Systems, Benchmark, Analysis and Evaluation of Security Systems, Distributed Access Control and Trust Management, Distributed Attack Systems and Mechanisms, Distributed Intrusion Detection/Prevention Systems, Denial-of-Service Attacks and Countermeasures, High Performance Security Systems, Identity Management and Authentication, Implementation, Deployment and Management of Security Systems, Intelligent Defense Systems, Internet and Network Forensics, Large-scale Attacks and Defense, RFID Security and Privacy, Security Architectures in Distributed Network Systems, Security for Critical Infrastructures, Security for P2P systems and Grid Systems, Security in E-Commerce, Security and Privacy in Wireless Networks, Secure Mobile Agents and Mobile Code, Security Protocols, Security Simulation and Tools, Security Theory and Tools, Standards and Assurance Methods, Trusted Computing, Viruses, Worms, and Other Malicious Code, World Wide Web Security, Novel and emerging secure architecture, Study of attack strategies, attack modeling, Case studies and analysis of actual attacks, Continuity of Operations during an attack, Key management, Trust management, Intrusion detection techniques, Intrusion response, alarm management, and correlation analysis, Study of tradeoffs between security and system performance, Intrusion tolerance systems, Secure protocols, Security in wireless networks (e.g. mesh networks, sensor networks, etc.), Cryptography and Secure Communications, Computer Forensics, Recovery and Healing, Security Visualization, Formal Methods in Security, Principles for Designing a Secure Computing System, Autonomic Security, Internet Security, Security in Health Care Systems, Security Solutions Using Reconfigurable Computing, Adaptive and Intelligent Defense Systems, Authentication and Access control, Denial of service attacks and countermeasures, Identity, Route and

Location Anonymity schemes, Intrusion detection and prevention techniques, Cryptography, encryption algorithms and Key management schemes, Secure routing schemes, Secure neighbor discovery and localization, Trust establishment and maintenance, Confidentiality and data integrity, Security architectures, deployments and solutions, Emerging threats to cloud-based services, Security model for new services, Cloud-aware web service security, Information hiding in Cloud Computing, Securing distributed data storage in cloud, Security, privacy and trust in mobile computing systems and applications, **Middleware security & Security features:** middleware software is an asset on
its own and has to be protected, interaction between security-specific and other middleware features, e.g., context-awareness, **Middleware-level security monitoring and measurement:** metrics and mechanisms for quantification and evaluation of security enforced by the middleware, **Security co-design:** trade-off and co-design between application-based and middleware-based security, **Policy-based management:** innovative support for policy-based definition and enforcement of security concerns, **Identification and authentication mechanisms:** Means to capture application specific constraints in defining and enforcing access control rules, **Middleware-oriented security patterns:** identification of patterns for sound, reusable security, **Security in aspect-based middleware:** mechanisms for isolating and enforcing security aspects, **Security in agent-based platforms:** protection for mobile code and platforms, Smart Devices: Biometrics, National ID cards, Embedded Systems Security and TPMs, RFID Systems Security, Smart Card Security, Pervasive Systems: Digital Rights Management (DRM) in pervasive environments, Intrusion Detection and Information Filtering, Localization Systems Security (Tracking of People and Goods), Mobile Commerce Security, Privacy Enhancing Technologies, Security Protocols (for Identification and Authentication, Confidentiality and Privacy, and Integrity), Ubiquitous Networks: Ad Hoc Networks Security, Delay-Tolerant Network Security, Domestic Network Security, Peer-to-Peer Networks Security, Security Issues in Mobile and Ubiquitous Networks, Security of GSM/GPRS/UMTS Systems, Sensor Networks Security, Vehicular Network Security, Wireless Communication Security: Bluetooth, NFC, WiFi, WiMAX, WiMedia, others

This Track will emphasize the design, implementation, management and applications of computer communications, networks and services. Topics of mostly theoretical nature are also welcome, provided there is clear practical potential in applying the results of such work.

Track B: Computer Science

Broadband wireless technologies: LTE, WiMAX, WiRAN, HSDPA, HSUPA, Resource allocation and interference management, Quality of service and scheduling methods, Capacity planning and dimensioning, Cross-layer design and Physical layer based issue, Interworking architecture and interoperability, Relay assisted and cooperative communications, Location and provisioning and mobility management, Call admission and flow/congestion control, Performance optimization, Channel capacity modeling and analysis, Middleware Issues: Event-based, publish/subscribe, and message-oriented middleware, Reconfigurable, adaptable, and reflective middleware approaches, Middleware solutions for reliability, fault tolerance, and quality-of-service, Scalability of middleware, Context-aware middleware, Autonomic and self-managing middleware, Evaluation techniques for middleware solutions, Formal methods and tools for designing, verifying, and evaluating, middleware, Software engineering techniques for middleware, Service oriented middleware, Agent-based middleware, Security middleware, Network Applications: Network-based automation, Cloud applications, Ubiquitous and pervasive applications, Collaborative applications, RFID and sensor network applications, Mobile applications, Smart home applications, Infrastructure monitoring and control applications, Remote health monitoring, GPS and location-based applications, Networked vehicles applications, Alert applications, Embeded Computer System, Advanced Control Systems, and Intelligent Control : Advanced control and measurement, computer and microprocessor-based control, signal processing, estimation and identification techniques, application specific IC's, nonlinear and adaptive control, optimal and robot control, intelligent control, evolutionary computing, and intelligent systems, instrumentation subject to critical conditions, automotive, marine and aero-space control and all other control applications, Intelligent Control System, Wiring/Wireless Sensor, Signal Control System. Sensors, Actuators and Systems Integration : Intelligent sensors and actuators, multisensor fusion, sensor array and multi-channel processing, micro/nano technology, microsensors and microactuators, instrumentation electronics, MEMS and system integration, wireless sensor, Network Sensor, Hybrid

Sensor, Distributed Sensor Networks. Signal and Image Processing : Digital signal processing theory, methods, DSP implementation, speech processing, image and multidimensional signal processing, Image analysis and processing, Image and Multimedia applications, Real-time multimedia signal processing, Computer vision, Emerging signal processing areas, Remote Sensing, Signal processing in education. Industrial Informatics: Industrial applications of neural networks, fuzzy algorithms, Neuro-Fuzzy application, bioInformatics, real-time computer control, real-time information systems, human-machine interfaces, CAD/CAM/CAT/CIM, virtual reality, industrial communications, flexible manufacturing systems, industrial automated process, Data Storage Management, Harddisk control, Supply Chain Management, Logistics applications, Power plant automation, Drives automation. Information Technology, Management of Information System : Management information systems, Information Management, Nursing information management, Information System, Information Technology and their application, Data retrieval, Data Base Management, Decision analysis methods, Information processing, Operations research, E-Business, E-Commerce, E-Government, Computer Business, Security and risk management, Medical imaging, Biotechnology, Bio-Medicine, Computer-based information systems in health care, Changing Access to Patient Information, Healthcare Management Information Technology. Communication/Computer Network, Transportation Application : On-board diagnostics, Active safety systems, Communication systems, Wireless technology, Communication application, Navigation and Guidance, Vision-based applications, Speech interface, Sensor fusion, Networking theory and technologies, Transportation information, Autonomous vehicle, Vehicle application of affective computing, Advance Computing technology and their application : Broadband and intelligent networks, Data Mining, Data fusion, Computational intelligence, Information and data security, Information indexing and retrieval, Information processing, Information systems and applications, Internet applications and performances, Knowledge based systems, Knowledge management, Software Engineering, Decision making, Mobile networks and services, Network management and services, Neural Network, Fuzzy logics, Neuro-Fuzzy, Expert approaches, Innovation Technology and Management : Innovation and product development, Emerging advances in business and its applications, Creativity in Internet management and retailing, B2B and B2C management, Electronic transceiver device for Retail Marketing Industries, Facilities planning and management, Innovative pervasive computing applications, Programming paradigms for pervasive systems, Software evolution and maintenance in pervasive systems, Middleware services and agent technologies, Adaptive, autonomic and context-aware computing, Mobile/Wireless computing systems and services in pervasive computing, Energy-efficient and green pervasive computing, Communication architectures for pervasive computing, Ad hoc networks for pervasive communications, Pervasive opportunistic communications and applications, Enabling technologies for pervasive systems (e.g., wireless BAN, PAN), Positioning and tracking technologies, Sensors and RFID in pervasive systems, Multimodal sensing and context for pervasive applications, Pervasive sensing, perception and semantic interpretation, Smart devices and intelligent environments, Trust, security and privacy issues in pervasive systems, User interfaces and interaction models, Virtual immersive communications, Wearable computers, Standards and interfaces for pervasive computing environments, Social and economic models for pervasive systems, Active and Programmable Networks, Ad Hoc & Sensor Network, Congestion and/or Flow Control, Content Distribution, Grid Networking, High-speed Network Architectures, Internet Services and Applications, Optical Networks, Mobile and Wireless Networks, Network Modeling and Simulation, Multicast, Multimedia Communications, Network Control and Management, Network Protocols, Network Performance, Network Measurement, Peer to Peer and Overlay Networks, Quality of Service and Quality of Experience, Ubiquitous Networks, Crosscutting Themes – Internet Technologies, Infrastructure, Services and Applications; Open Source Tools, Open Models and Architectures; Security, Privacy and Trust; Navigation Systems, Location Based Services; Social Networks and Online Communities; ICT Convergence, Digital Economy and Digital Divide, Neural Networks, Pattern Recognition, Computer Vision, Advanced Computing Architectures and New Programming Models, Visualization and Virtual Reality as Applied to Computational Science, Computer Architecture and Embedded Systems, Technology in Education, Theoretical Computer Science, Computing Ethics, Computing Practices & Applications

Authors are invited to submit papers through e-mail ijcsiseditor@gmail.com. Submissions must be original and should not have been published previously or be under consideration for publication while being evaluated by IJCSIS. Before submission authors should carefully read over the journal's Author Guidelines, which are located at http://sites.google.com/site/ijcsis/authors-notes .

ISSN 1947 5500
http://sites.google.com/site/ijcsis/