# SOM BASED ASSURANCE ASSESSMENT FOR INFORMATION SECURITY MANAGEMENT SYSTEM IN ORGANIZATION

Tubagus Mohammad Akhriza*

***Abstract:*** *Information Security Management System (ISMS) has a lot of standards such as ISO/IEC 27001. In order to assess the quality assurance of ISMS, an organization should design and implement some mechanism. I propose a mechanism for quality assurance assessment using Kohonen's Self Organizing Map (SOM). SOM is a means for automatically arranging high-dimensional statistical data so that alike inputs are in general mapped close to each other. The resulting map avails itself readily to visualization, and thus the distance relations between different data items can be illustrated in an intuitive manner. This proposal shows that the proposed framework is able to be conducted and implemented. Through this map, the organization will be able to assess how far the organization's information security quality has gap with the standard. Not only ISMS ISO's standard, but also another standard such as Indonesian National Board of Accreditation (BAN-PT).*

***Keywords:*** *SOM, Information Security Management System, Quality Assurance.*

## INFORMATION SECURITY MANAGEMENT SYSTEM AND INFORMATION VISUALIZATION.

Information has become one of important assets for organization. In this information era, information is delivered likes no border between the society communities. Considering this condition, organization needs to determine strategy to secure their information in order to minimize the risk of information. This condition arise a lot researches about information security management system. The international organization for standardization (ISO) also publishes some standards for ISMS. Some approaches also have been developed to assess the quality of information security management system in all aspects. Meanwhile, in information retrieval and information visualization field of research, also have been developed a lot of methods. Information visualization tools can be used as assessment or evaluation tools for multidimensional data and information because they have ability to display all information at once on graphical output. This phenomenon leads the author to implement information visualization method in information security management system quality assessment.

## LIMITATION

This paper only discuss about a little part of ISMS ISO 27001 i.e. control of

* Author is a lecturer of STMIK Pradnya Paramita Malang and studying for PhD Degree in The School of Information Security Engineering, Shanghai Jiao Tong University, P.R. of China.

records and a proposal about building a framework to assess quality of the control of records management using SOM.

## PAPER'S ORGANIZATION

This paper is organized as follows:

- Part 1 – Introduction, contains the author's motivation and limitation of paper
- Part 2 – ISMS ISO Standard 27001:2005
- Part 3 – Information Classification
- Part 4 – Measuring the Control of Record
- Part 5 – Kohonen's SOM
- Part 6 – Proposed Framework
- Part 7 – Conclusion.

## ISMS ISO STANDARD 27001:2005

### Brief Introduction

According to [2], ISO (the International Organization for Standardization) and IEC (the International Electro-technical Commission) form the specialized system for worldwide standardization. ISO/IEC 27001 was prepared by Joint Technical Committee ISO/IEC JTC 1, Information technology, Subcommittee SC 27, IT Security techniques.

According to [2] in the part of term and definitions, mentioned that Information Security Management System (ISMS) is that part of the overall management system, based on a business risk approach, to establish, implement, operate, monitor, review, maintain and improve information security. ISMS have to be able to assure that the data is available, approved and signed by legal parts or persons.

### 2.2. Control of Records

One part in the ISO 27001 describe about the control of records. In [2] mentioned that records shall be established and maintained to provide evidence of conformity to requirements and the effective operation of the ISMS. They shall be protected and controlled. The ISMS shall take account of any relevant legal or regulatory requirements and contractual obligations.

Records shall remain legible, readily identifiable and retrievable. The controls needed for the identification, storage, protection, retrieval, retention time and disposition of records shall be documented and implemented.

## 3. INFORMATION CLASSIFICATION

In order to describe the standard of control of records in ISO 27001, we can make classification of information. We can determine the information classification by two properties: availability classification and sensitivity classification.

## Availability Classification

Information of different types needs to be secured in different ways. Therefore a classification system is needed, whereby information is classified, a policy is laid down on how to handle information according to its class and security mechanisms are enforced on systems handling information accordingly

Availability classification is arranged in order to determine the measure of availability document measurement. In [8] we can find an example of availability classification as showed in table 1. The higher rank the more available information.

Table 1: Availability Classification

| Class | Availability Rank | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| Maximum allowed Server *downtime*, per event | 1 Week | 1 Day | 1 Hour | 1 Hour |
| On which Days? | Mon-Fri | Mon-Fri | Mon-Fri | 7 Days |
| During what hours? | | | 07:00-18:00 | 24h |
| Expected availability percentage | 80% | 95% | 99.5% | 99.9% |
| expected max. downtime | 1 day/week | 2 hours/Week | 20min./Week | 12min./month |

## Sensitivity Classification

A classification system is proposed which class information / processes into four levels. For example, the lowest 1 is the least sensitive and the highest 4 is for the most important information / processes.

The basic concept of the data or information for example shown as follow:

All data has an owner.

- The data or process owner must classify the information into one of the security levels- depending on legal obligations, costs, corporate into policy and business needs.
- If the owner is not sure at what level data should be classified, use level 3.
- The owner must declare who is allowed access to the data.

- The owner is responsible for this data and must secure it or have it secured (e.g. via a security administrator) according to it's classification.
- All documents should be classified and the classification level should be written on at least the title page.

The Information Classification Policy standardizes how information is categorized by its level of sensitivity. There are many ways to classify information. Categorizations should be based upon business use [3]. The military has several ways of categorizing information. The more general categories are "Unclassified" and "Classified." Additional categorizations are "Public," "For Official Use Only," "Sensitive," "Secret," and "Top Secret." Specific rules are defined in the Information Classification Policy for how these types of information should be distributed, retained, protected, and destroyed. While "For Official Use Only" information can sit openly on a desk, it is not typically released to the general public and it should be shredded when it becomes obsolete.

Organization needs to classify the information according the organization structure. For example, anything not "Public" is treated as "Sensitive," and the rules for "Sensitive" information could be the same as those for the highest level of classification as defined by the organization. The Information Disclosure, Dissemination, or "Media" Policy defines how information will be disseminated. Security incidences are extremely sensitive and can impact the success of the business.

Take the example of a major bank that has had an incident of unauthorized access to its client accounts. When knowledge of the incident becomes public, the bank is going to have a significant loss in clients due to trust, possible negative shareholder activity, and repercussions from its board of directors. The Information Disclosure Policy defines a central point of contact for distributing information to the public or interfacing with the media.

We can divide the sensitivity of information in to 4 classes [7],[8]:

- Class 1: Public / non classified information. Data on these systems could be made public without any implications for the company (i.e. the data is not confidential). Data integrity is not vital. Loss of service due to malicious attacks is an acceptable danger. Examples: Test services without confidential data, certain public information services, products brochures widely distributed, data available in the public domain anyway.
- Class 2: Internal information. External access to this data is to be prevented, but

should this data become public, the consequences are not critical (e.g. the company may be publicly embarrassed). Internal access is selective. Data integrity is important but not vital. Examples of this type of data are found in development groups (where no live data is present), certain production public services, certain Customer Data, "normal" working documents and project/meeting protocols, Telephone books.

- Class 3: Confidential information. Data in this class is confidential within the company and protected from external access. If such data were to be accessed by unauthorized persons, it could influence the company's operational effectiveness, cause an important financial loss, provide a significant gain to a competitor or cause a major drop in customer confidence. Data integrity is vital. Examples: Datacenters normally maintain this level of security: salaries, Personnel data, Accounting data, passwords, information on corporate security weaknesses, very confidential customer data and confidential contracts.
- Class 4: Secret information. Unauthorized external or internal access to this data would be critical to the company. Data integrity is vital. The number of people with access to this data should be very small. Very strict rules must be adhered to in the usage of this data. Examples: Military data, secret contracts.

As an example, we can determine the sensitivity of information and access control list (ACL) as shown in table 2.

Table 2: Information Classification and Access Control List

| Documents | User Level | | | |
|---|---|---|---|---|
| | Managers | Staffs | Customers | Public |
| Phone number book | R | RW | | |
| Vulnerability disclosure | RW | | | |
| Product brochure | RWEM | RW | R | R |
| Customer's PIN | | | R | |

R = Read, W = Write, E=Erase, M=Modify

## MEASURING THE CONTROL OF RECORDS

As mentioned in Part 2.2 that records shall remain legible, readily identifiable and retrievable. The controls needed for the identification, storage, protection, retrieval, retention time and disposition of records shall be documented and implemented.

From that statement, at least we can conduct the property of records and the property of the control:

- The properties of records are: legible, identifiable and retrievable
- The properties of controls are: have identification, storage, protection, retrieval, retention time and disposition.

One of the challenge is those properties are calculated from large number of parameters of assessment. It means they have multidimensional parameter and we want to reduce the multidimensional become 2 or 3 dimensional so we can represent them on the map of 2 or 3 dimensions.

At present there are several methods of dimensionality reduction as follows [9]: Principal Component Analysis (PCA), Multi-dimensional Scaling (MDS), Self-Organizing Map (SOM), Hierarchical Clustering, and Pathfinder Network Scaling. The methods used more are MDS, SOM. In the visual space constructed by MDS, the distance between data points is in proportion to the similarity between various parameters in the high-dimensional vector space. The high-dimensional data is mapped to 2-D visual region in the map by using SOM, where the color and the region size represent the density of the parameters.

We can find a lot of SOM implementation in visualizing information. One of the researches in [10] is proposed to solve network administration problem by finding the solution in the cluster of articles related to the network administration problem solving. The authors present a methodology for defining the vocabulary and preliminary results for assessing the quality of expert-proposed modifications to the vocabulary. They obtain vocabulary-derived document classes from a self-organizing map and assess vocabulary quality using the quality of classification into these classes.

## KOHONEN'S SOM

The self-organizing map (SOM) [4], [5],[6] is a means for automatically arranging high-dimensional statistical data so that alike inputs are in general mapped close to each other. The resulting map avails itself readily to visualization, and thus the distance relations between different data items can be illustrated in an intuitive manner.

In [1] we can get a very clear online tutorial about SOM. A common example used to help teach the principals behind SOMs is the mapping of colors from their

three dimensional components - red, green and blue, into two dimensions. Figure 1 shows an example of a SOM trained to recognize the eight different colors shown on the right. The colors have been presented to the network as 3D vectors - one dimension for each of the color components - and the network has learnt to represent them in the 2D. Notice that in addition to clustering the colors into distinct regions, regions of similar properties are usually found adjacent to each other. This feature of Kohonen maps is often put to good use as you will discover later.

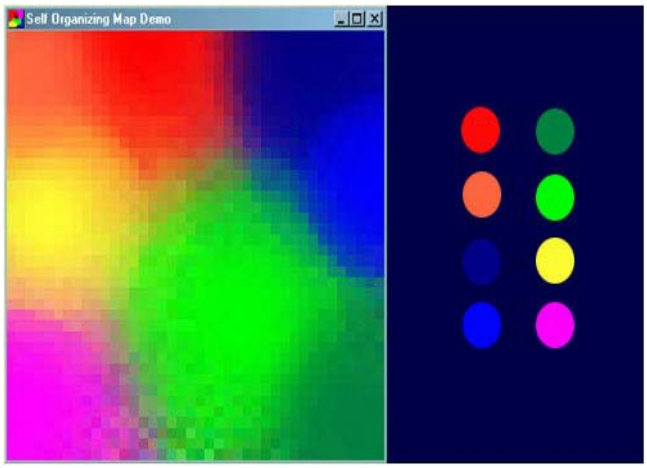


Figure 1: SOM result for color clustering (left) and the training data set of colors (right)

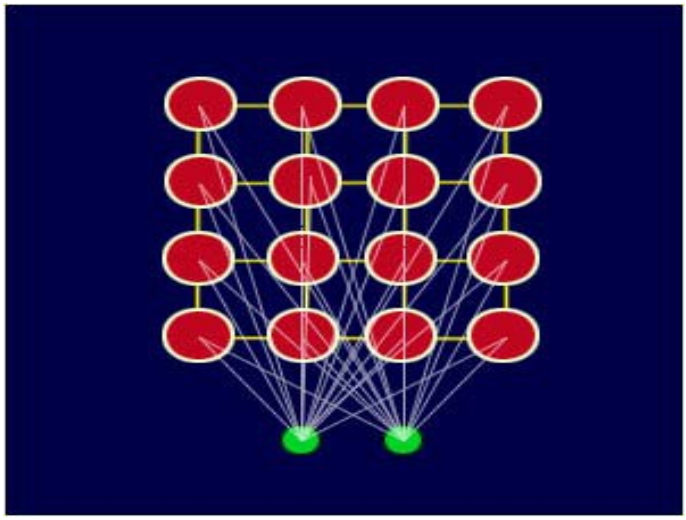

Figure 2: A Simple Kohonen Network

For the color clustering, the network is created from a 2D lattice of 'nodes', each of which is fully connected to the input layer. Figure 2 shows a very small Kohonen network of 4 X 4 nodes connected to the input layer (shown in green) representing a two dimensional vector.

Each node has a specific topological

position (an x, y coordinate in the lattice) and contains a vector of weights of the same dimension as the input vectors. That is to say, if the training data consists of vectors, V, of n dimensions:

$$V_1, V_2, V_3...V_n$$

Then each node will contain a corresponding weight vector W, of n dimensions:

$$W_1, W_2, W_3...W_n$$

The lines connecting the nodes in Figure 2 are only there to represent adjacency and do not signify a connection as normally indicated when discussing a neural network. There are no lateral connections between nodes within the lattice.

A SOM does not need a target output to be specified unlike many other types of network. Instead, where the node weights match the input vector, that area of the lattice is selectively optimized to more closely resemble the data for the class the input vector is a member of. From an initial distribution of random weights, and over many iterations, the SOM eventually settles into a map of stable zones. Each zone is effectively a feature classifier, so you can think of the graphical output as a type of feature map of the input space. If you take another look at the trained network shown in figure 1, the blocks of similar color represent the individual zones. Any new, previously unseen input vectors presented to the network will stimulate nodes in the zone with similar weight vectors.

Training occurs in several steps and over many iteration:

1. Each node's weights are initialized.
2. A vector is chosen at random from the set of training data and presented to the lattice.
3. Every node is examined to calculate which one's weights are most like the input vector. The winning node is commonly known as the Best Matching Unit (BMU).
4. The radius of the neighborhood of the BMU is now calculated. This is a value that starts large, typically set to the 'radius' of the lattice, but diminishes each time-step. Any nodes found within this radius are deemed to be inside the BMU's neighborhood.
5. Each neighboring node's (the nodes found in step 4) weights are adjusted to make them more like the input vector. The closer a node is to the BMU, the more its weights get altered.
6. Repeat step 2 for N iterations.

To determine the best matching unit, one method is to iterate through all the nodes and calculate the Euclidean distance between each node's weight vector and the current input vector. The node with a weight vector closest to the input vector is

tagged as the BMU.

The Euclidean distance is given as:

$$Dist = \sqrt{\sum_{i=0}^{i=n} (V_i - W_i)^2} \quad (1)$$

where $\boldsymbol{V}$ is the current input vector and $\boldsymbol{W}$ is the node's weight vector

As an example, to calculate the distance between the vector for the color red (1, 0, 0) with an arbitrary weight vector (0.1, 0.4, 0.5)

Distance = sqrt( $(1 - 0.1)^2 + (0 - 0.4)^2 +$
$(0 - 0.5)^2$ )
= sqrt( $(0.9)^2 + (-0.4)^2 + (-0.5)^2$ )
= sqrt( $0.81 + 0.16 + 0.25$ )
= sqrt(1.22)
Distance = 1.106

For each iteration after the BMU has been determined, the next step is to calculate which of the other nodes are within the BMU's neighborhood. All these nodes will have their weight vectors altered in the next step. First we calculate what the radius of the neighborhood should be and then it's a simple application of Pythagoras to determine if each node is within the radial distance or not.

Figure 3 shows an example of the size of a typical neighborhood close to the commencement of training.

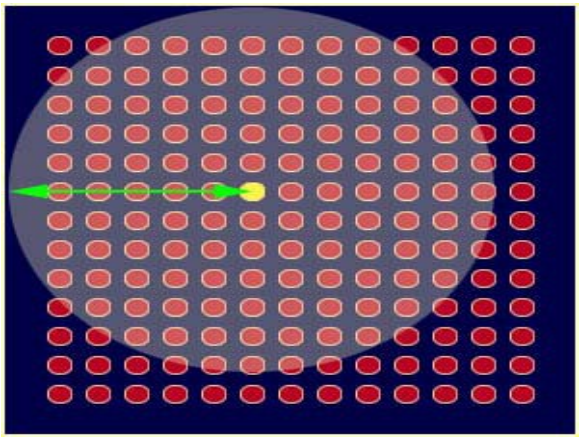

Figure 3: The BMU's neighbourhood.

We can see that the neighborhood shown above is centered around the BMU (colored yellow/the center of the circle) and encompasses most of the other nodes. The green arrow shows the radius

A unique feature of the Kohonen learning algorithm is that the area of the neighborhood shrinks over time. This is accomplished by making the radius of the neighbourhood shrink over time To do this we can use the exponential decay function:

$$\sigma(t) = \sigma_0 \exp\left(-\frac{t}{\lambda}\right) \quad t = 1,2,3,... \tag{2}$$

where the Greek letter sigma, $\sigma_0$, denotes the width of the lattice at time $t_0$ and the Greek letter lambda, $\lambda$, denotes a time constant. $t$ is the current time-step (iteration of the loop)

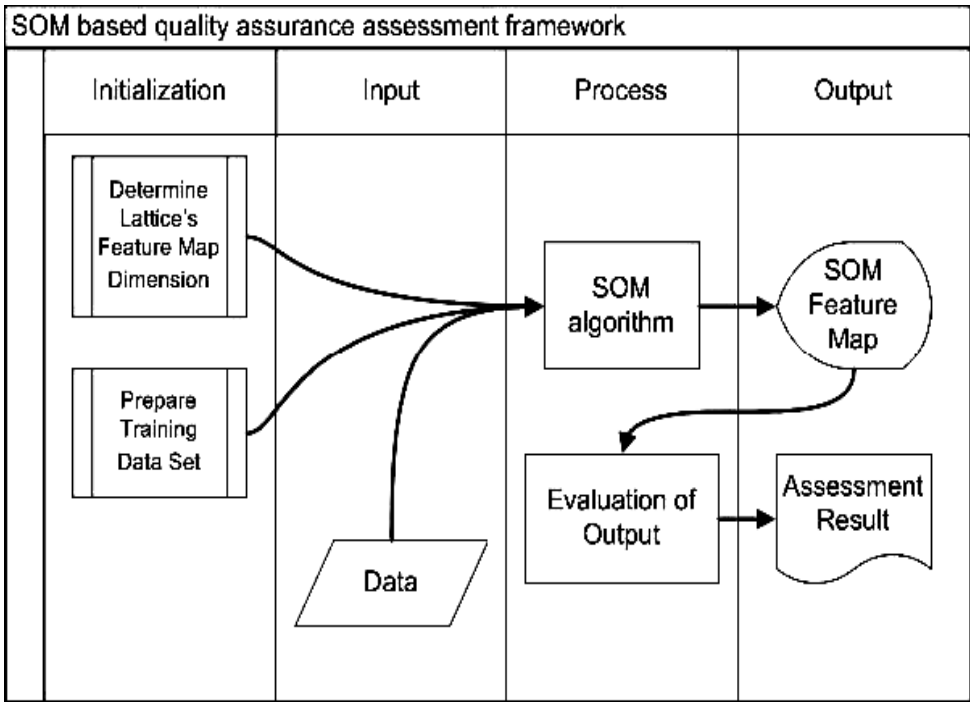


Figure 4: Proposed Framework

## PROPOSED FRAMEWORK

### The Framework

The framework of this approach is shown in figure 4 explanation of framework is as follows.

### Lattice Dimension

Lattice dimension or feature map dimension or output map dimension is the size of the output map. For example 10 x 10 means that the output map will contain at most 100 data.

### Training Data Set

As we can learn from SOM algorithm's characteristic, the data that we want to assess and the data in the training set must have same vector dimension. In this section we will determine the vector of the data according to the control of documents standard.

Based on the sample of training data set of colors, we know that a color has 3 dimensions i.e. (red, green, blue). In normalized form, the red has vector of (1, 0, 0), green has (0, 1, 0) and blue (0, 0, 1). 1 means the intensity of that vector color is the highest and 0 means the lowest. So if a color has vector (0, 0, 0) then it means the color is black, and also vector (1, 1, 1) means white. We can consider that if find a vector of (1,0,1) then the color is purple.

Considering the analogy of color sense, we may determine the vector of record and control as follows:

1. The record is a 3 dimensions vectors of legible, identifiable and retrievable
2. The control is a 6 dimensions vectors of has identification, storage, protection, retrieval, retention time and disposition
3. Or if we want to join the record and control become the control of record then we have 9 dimensions vector of all of their properties.

The most important thing is we have to determine the parameter of each property, for example what is so called the eligible, identifiable and retrievable record? If want to set the highest value is 1 and the lowest is 0, then we can determine that the best quality of record has vector (legible=1, identifiable=1, retrievable=1). This approach also can be done for 'the control'.

### Data

Data is the material that we collect from organization's work-unit. This data originally contents all about performance of work-unit in order to manage the information security. Or we can also say that this data has multidimensional. But since this data will be assessed using the SOM, we have to change the dimension of the data so they have the same vector dimension with the training data set.

At the initial step of SOM algorithm,

this data will be spread out on to the lattice. After that by evaluating each data with each training data set, those data will be clustered automatically according to the similarity of the vector between the data and the training data set.

## SOM Feature Map

SOM feature map or output map is an m x m dimension lattice that we set up in the initialization step. It could be rectangular matrix or unified distance matrix (u-matrix) as shown as figure 5.

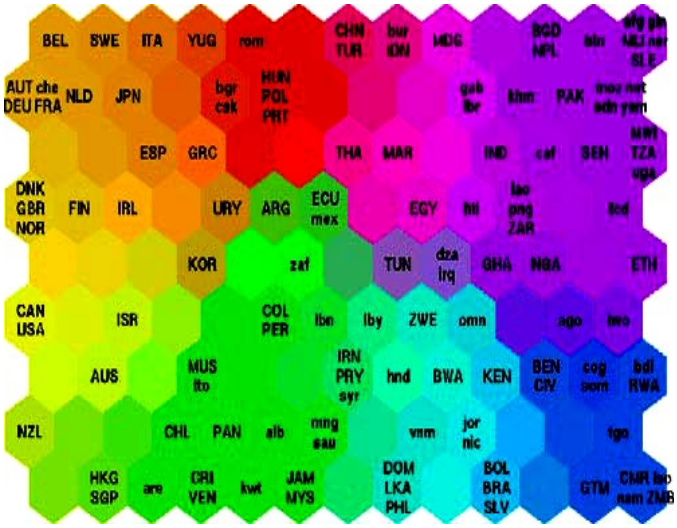


Figure 5. A 9 x 13 Dimension U-Matrix

## Evaluation of Output

This step contains two processes:

1. The experiment to evaluate feature map of records, control and the control of record.
2. The assessment of quality of the control of records

1

## Expected Result

The expected of the result is the SOM's feature map of the quality of record, control and the control of record. Using this visualization tools we hope the map will help the organization to assess the quality of the control of records as part of Information Security Management System (ISMS).

## CONCLUSION

Considering the framework and the approach, this proposal is possible to be conducted and implemented.

This idea is not only could be implemented for ISMS ISO's standard, but also could be implemented in another standard. In Indonesia we know some standards for Academic Quality Assurance such as from The National Board of Accreditation for Higher Education (Badan Akreditasi Nasional – Perguruan Tinggi/ BAN-PT).

As long as we could determine the numeric value of each parameter, then we could run this SOM mechanism to assess how far is our higher education institution's annually targets with the BAN-PT's standard.

## REFERENCE


[1] Anonymous. Without year. Kohonen's Self Organizing Feature Maps. http://www.ai-junkie.com/ann/som/som1.html.

[2] ISO. 2005. Information technology–Security techniques–Information security management systems–Requirements. International Standard ISO/IEC 27001. First Edition 2005-10-15.

[3] Killmeyer, J., 2006. Information Security Architecture. An Integrated Approach to Security in the Organization Second Edition. Auerbach Publications Taylor & Francis Group 6000 Broken Sound Parkway NW,

[4] Kohonen. 1982. Self-organizing formation of topologically correct feature maps. Biological Cybernetics, 43(1):59-69

[5] Kohonen, T. 1995. Self-Organizing Maps. Springer. Berlin. Heidelberg. 2nd extended ed. 1997

[6] Kohonen, T., Hynninen, J., Kangas, J., and Laaksonen, J. 1996. SOM_PAK: The Self-Organizing Map program package. Report A31, Helsinki University of Technology, Laboratory of Computer and Information Science.

[7] Pfleeger, C.P. and Pfleeger , S.L. 2006. Security in Computing, 4th Edition. Prentice Hall PTR; 4th edition (October 23, 2006)

[8] Sean Boran. 2003. IT Security Cookbook. http://www.boran.com/security/

[9] Fangfang Zhang and Shaolin Deng. 2008. Studies on the Visualization for Web Information Retrieval. IEEE Computer Society.

[10] George, A., Makanju, A., Heywood, A.N.Z., and Milios, E.E. 2008. Information Retrieval in Network Administration. IEEE Computer Society.