Design and Development of a Security Evaluation Platform Based on International Standards

Yuji Takahashi and Yoshimi Teshigawara

Graduate School of Engineering, Soka University, Japan
{e08d5203, teshiga}@soka.ac.jp

Abstract - To obtain security attestation, organizations evaluate security products by using systems based on international standards. However, they currently must use individual systems corresponding to different versions of these standards. Therefore, we have been studying a platform that enables evaluation for different standard contents and evaluation targets by focusing on changes of the standards used as evaluation criteria. We developed and implemented the platform taking into consideration the hierarchical structure and reference relations of the standards. The platform provides functions such as a reference-related arrangement of the whole standard, the display of a reference tree, and score calculation. In addition, in order to produce the pertinent information for data conversion, we calculated the similarity between two standards. Experimental evaluation shows that covering all items and avoidance of human error can be achieved by supplementing technical knowledge and by utilizing visual effects. The validity of the platform is also confirmed.

Keywords: Security management, Information security, International standard, ISO/IEC 27000

1 BACKGROUND AND PURPOSE OF RESEARCH

In recent years, the scope of security management is expanding from self-defense for protecting the assets of an organization to preventing becoming the target of attackers who cause damage to the organization. As a result, it has become important to have the status of the implementation of safety and security measures assessed by an external agency [1]. There is a specific standard for such assessment, called ISO/IEC 27001 and the number of organizations that are being accredited by this standard is continuously increasing. By June 2012 more than 7,000 companies had been accredited worldwide, more than 4,000 of them in Japan [2].

For most of the security certifications, standards such as ISO/IEC 27001, ISO/IEC 27002, and JIS Q 15001 are taken as references and organizations are accredited by satisfying all the items that are described in those standards. In addition, security assessment systems are used to validate the achievement of criteria in the certification process [3]. However, the items of the standard are frequently changed as time passes. Compared with other standards, security related standards are changed more frequently because they are not tested precisely; user comments are taken into consideration and changes are made accordingly. In addition, because the certification process differs depending on the size of the organization and other factors, the criteria for assessment also differ. If the organization and the objective of the assessment change, changes such as revision of the standard will create a situation where a new system has to be created for redoing each certification using individual tools or personnel. Hence considerable time, personnel and money are required and this leads to problems that have huge personnel and monetary impacts on company activities. The need for a general assessment tool, allowing changes in the organization and the purpose of assessment, in place of individual security assessment tools, has been increasing.

We have been studying a security assessment platform that enables the realization of particular security assessments by replacing only raw data (hereinafter referred to as fundamental data) that address the fundamental standard without depending on the target standard [4]. In this platform, by focusing on the hierarchical structure of sentences, which is a characteristic structure of standards documents, the items of the standard as well as statements indicating the detailed conditions and references to other items (hereinafter reference relations) were organized in a hierarchical structure. Security assessment needs to be done without depending on the type of reference standard, and a method of estimating the assessment level without depending on the type of standard is required. In the platform the assessment level is estimated using the hierarchical structure and reference relations, so we have been studying a platform that aims to be suitable for achieving security requirements. In particular, we have developed an appropriate platform system and registered the data for the ISO/IEC 27000 series [5]. In this study, security assessment is conducted by changing the impact of assessment with respect to each component of the reference tree as described below. On the basis of experimental results for security assessment methods that consider the distance in the reference trees as well as the relation of each item with assessment items, we found that changing impact is effective. For considering the relation of each item with assessment items, we proposed various methods of estimating impact and experimented using these methods [5] [6] [7]. Thus, for those users who do not have a deep knowledge of attestation, we experimented with changing the sample providing function and the data migration function, using relevant information based on past cases, in order to support countermeasure selection and implementation, and we confirmed the validity of the proposed platform [7] [8]. We found that the effectiveness of the data migration function can be increased by interlocking with the sample function

[8]. In this paper, we evaluate the results of the experiments done for each function separately and in combination.

2 ANALYSIS AND UTILIZATION OF STANDARDS

2.1 Relevant standards

In this paper, the experiments and verifications are performed mainly using the security standard data that have been summarized as the ISO/IEC 27000 series. This has adapted the concept of the Plan-Do-Check-Act (PDCA) cycle, which is widely used in the standards of security management and is represented in information security management systems (ISMSs).

This security assessment platform is intended to be used not in a single phase of the PDCA cycle but in every phase where the platform is applicable. If it is applied in the Plan stage, the loopholes in the countermeasures can be checked by entering the results of the present data analysis. In the Do stage when it is recognized that enforcing countermeasures does not cover the planning item, the loophole can be verified in its entirety by means of checking those items. In the Check stage, the functionality of each countermeasure can be checked according to the plan made in the countermeasure enforcement stage. The loopholes can be checked by summing those changes in the corresponding conditions that match the conditions in practice. In the Act stage, as in the Plan stage, the loopholes of the corresponding countermeasures that were re-defined can be checked.

2.1.1. ISO/IEC 27000-series

The ISO/IEC 27000-series is an information security standard family, established by the collaboration between the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC). This series is broad in scope, covering privacy, confidentiality and information technology security issues. Therefore, it is applicable to organizations of all sizes and types.

To obtain security attestation in this series, organizations first assess their information security risks and then implement appropriate information security controls according to their needs. Given the dynamic nature of information security, the ISMS concept incorporates continuous feedback and improvement activities based on the PDCA cycle. As of June 2011, 10 standards of ISO/IEC 27000 had been developed and many other standards are now under development [9]. ISO/IEC 27000 is a standard reference in many areas and it shows the importance of the PDCA cycle to ISMSs.

2.1.2. ISO/IEC 27001

The objective of ISO/IEC 27001 is to provide a model for the establishment, implementation, operation, monitoring, review, maintenance and improvement of an ISMS [10]. In addition, the contents shown in each item of this standard in the operational manual created during the process of ISMS attestation, corresponds to the security requirements. It should cover all the items including those that specify what is outside the scope of the attestation. During the inspection for ISMS attestation, the security countermeasures corresponding to each item of this manual will be subject to inspection.

2.2 Standard configuration

Generally the body in the relevant standard has often been described in a hierarchical structure of three phases: 'Chapter', 'Section' and 'Item'. In a 'Chapter', the assessment targets are roughly classified. In a 'Section' the assessment targets are described in detail and in an 'Item' the contents are further described in more detail.

However, there are many individual items which are not only described as separate items but also as conditions or supplementary matters that refer to other items. For instance, Section 7.1 of ISO/IEC 27001 contains a reference to Item 4.3.3 and this relationship is expressed in the reference tree used in this study, as shown in Fig.1.

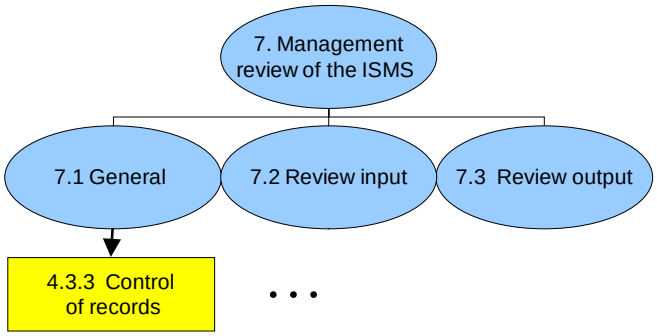

Figure 1: Reference-related example of ISO/IEC 27001

2.3 The problems of covering items related to countermeasures and their solution

In security attestation, the criteria should be comprehensively covered. Depending on the framework of each chapter of the configuration, the required policy decisions, such as implementation of countermeasures and acceptance of the risk, will be made. At that time, since there is a need for the comprehensive cover of the standard for each chapter, it is necessary to capture precisely the hierarchical structure of each chapter and reference relations for each item.

However, in all standards, not just ISO/IEC 27001, there are many references and there is a wide variety of content (items) which should be covered. Hence, understanding all of them precisely and choosing comprehensive measures becomes difficult. Therefore, it is desirable to manage collectively all the items covered in each chapter. To solve these issues, we propose a platform that can collectively manage all the items to be covered by using the hierarchical structure and the reference relations. Since the hierarchical structure and reference relations that are described in the

platform describe information from the standard with similar characteristics, the platform can cope if there is a change in the standard or even if the standard is a different one.

3 OVERVIEW OF THE PLATFORM

3.1 Structure of the platform

This platform consists of three parts namely, the data input unit, the data management unit and the score calculation unit. The configuration of the platform is shown in Fig.2. In the data input unit, the fundamental data of the standard, structural information, reference information, countermeasure information and other relevant information are entered. Initially the input of countermeasure information can be based on sample information created by the data management unit. Based on these fundamental data and the structural information, the data management unit organizes the data, develops the reference relations by using the reference information and configures the reference tree. In the score calculation unit, the calculated assessment values (score data) are managed. Also, based on the countermeasure information or other relevant information that has been input, the sample data are generated. In the score calculation unit, based on the reference information stored in the reference tree and the information about the registered countermeasures, the assessment value is calculated and the calculated data are passed to the data management unit.

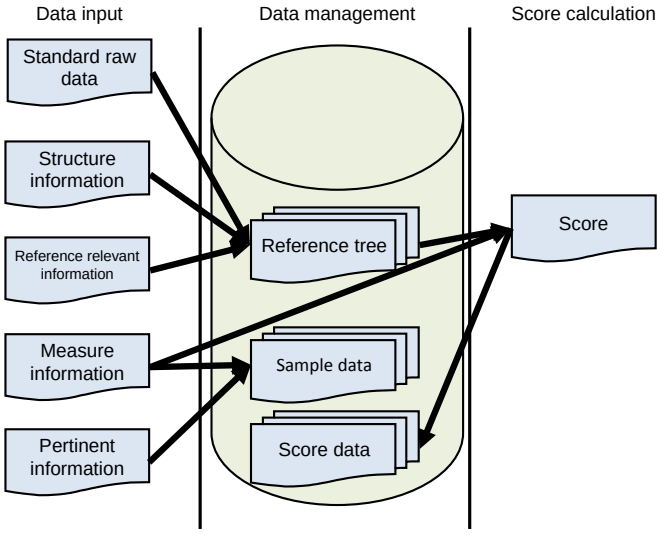

Figure 2: Structure of proposed platform

3.2 Behavior of the platform

First, the standard's fundamental data from the data input unit are stored. Then, the structural information based on the hierarchical structure described in section 2.3 is stored along with previously registered data. Subsequently, the hierarchy-based information and the direct reference information (hereinafter referred to as direct references) that are described in the standard document are registered. The registered multiple criteria (standards) are related to each other and, if relevant information is provided showing the requirement of measures, in terms of which items of each criterion are related to other items, that information is also registered. After data registration is completed, the registered data are delivered to the data management unit and then migrated to the next operation.

In this platform, the hierarchy is defined using levels. Chapters are defined as level '1' and the following stages as level '2' and so on. Level 'm' is assumed to refer directly to the items of level 'm+1'. In this study, this type of hierarchical structure is also defined as a part of the reference relations.

The basic tree is configured with an item that has a direct reference as the root (hereinafter referred to as the parent reference) and the described items that should be referenced (hereinafter referred to as a reference) are the leaves of the tree. If the leaf of a basic tree becomes the root of another basic tree, a new tree combining the part of the leaf of the former tree with the root of the latter tree is configured. During configuration, a leaf may have the same item as a reference as the root of the tree. If this repeated reference relation has multiple references at multiple locations with the same field as a reference, it will cause a reference loop to occur when the tree is configured. When these references occur, the part that overlaps is designated as the leaf and the configuration of the tree is continued. Thus, binding of the tree is continued until it becomes impossible to bind further and the largest tree becomes the reference tree.

In a reference tree, the relation between the items is expressed as a distance. The distance of those that are referenced directly is 1 and for each iteration of the following references the distance between the items increases gradually.

Subsequently, a standard for security attestation using that reference tree is created in the score calculation unit. The criterion is intended to provide an assessment value for the entire reference tree. In fact, in the data input unit, information about the countermeasure implementation, based on the information of the reference tree, countermeasures in past projects included in the sample data and the compliance status of each item in the standard, is suggested.

Based on the countermeasure information and the reference tree information that has been input, an assessment value is calculated. In addition, if the sample data are set in the data management unit during that time, countermeasures and the supporting data within the compliance status information of each item will be stored as sample data. After that, when the compliance status of the corresponding countermeasure is input by another user, the sample data can be input referring to the sample data that have already been provided.

If relevant information on other criteria is referred to when the assessment is done under new criteria, by means of the data migration function in the data management unit, sample data are generated based on the compliance status data of the underlying criteria and the data can be input while browsing.

3.3 Features of the platform

In this platform, when there is a change in the standard, the information in the data input unit is updated. After updating the information, the reference tree will be automatically reconfigured in the data management unit. In the score calculation unit, reassessment and the recalculation of the score can be done in accordance with the changed contents of the standard.

In addition, the relationships between the items can be visualized by configuring the reference tree. Choosing the countermeasures while checking the reference tree can help to set the effective countermeasures. In the sample data display function, managers who may not have sufficient expertise can share information. In the data migration function, during the reassessment process, the sample data that can be used as reference can be generated without any extra effort.

3.4 System configuration of the platform

This platform has been developed in Visual Basic, and various experiments have been performed so far. First, the entire platform is configured as a single program. The program is composed of independent subprograms: a subprogram that composes information for configuring the reference tree, after registering the criteria, hierarchical structure information and reference relation information; a subprogram that displays the reference tree; a subprogram that organizes the status of the countermeasures; and a subprogram that performs assessment value calculation. These subprograms ensure smooth running of the system by running in the background. For instance, when the data are first registered or when any change is made to the data, changes are made to the reference relations of all the criteria in the background and so, even during the process of making changes, the history of the data can be viewed. In addition, the body of the platform can always be run by operating the time consuming subprograms, such as displaying the reference tree and changing the status of the countermeasures, as independent programs during the process.

In addition, as the assessment value calculation is a separate subprogram, it can be easily changed to a new method. This is useful when introducing or testing multiple methods of assessment value calculation. Similarly, in the display function of the reference tree, instead of replacing the entire program to meet the user's demands, a display program that matches the user's preferences can be easily introduced.

4 METHOD OF CALCULATING IMPACT OF EACH COMPONENT OF THE REFERENCE TREE

In this study, which focuses on the number of items in the reference tree and the distances between them, the value of the assessment can be compared using the security assessment method that changes the impact of each component. In addition when items from other chapters are referred to in the reference tree, it is possible to determine the impact on the calculation results of those items due to the change in the calculation method.

There are four methods tested so far. Method 1 focuses only on the component number. The numbers of existing measures, measures in progress and measures yet to be implemented, in the reference tree which is the root of the estimated item is called '*n*'. The *i*th component is given the value x_i, where x_i is equal to 1 if the estimation item is applicable and is equal to 0 otherwise. The evaluation value $Score_1$ is given by

$$Score_1 = \frac{\sum_{i=1}^{n} x_i}{n} \tag{1}$$

Method 2 is an estimation method depending on the maximum distance. For the *i*th component of the reference tree which is the root of the evaluation item, the distance is d_i, and the maximum distance is d_{max}. As for Method 1 the component number is '*n*' and the *i*th component has the value x_i, which is equal to 1 if the estimation item is applicable and equal to 0 if it is not applicable. The degree of impact of the *i*th item is taken as $d_{max}–d_i,+1$. The evaluation value $Score_2$ is given by

$$Score_2 = \frac{\sum_{i=1}^{n} \{x_i (d_{\max} - d_i + 1)\}}{\sum_{i=1}^{n} (d_{\max} - d_i + 1)} \tag{2}$$

In this method, though there is change in the impact based on the maximum distance in the reference tree, depending on the distance, the impact of the degree of assessment is determined in monotonically decreasing form. Characteristically, the impact of each item on the assessed item falls slowly.

Method 3 uses the reciprocal of the distance. The assessment value $Score_3$ is given as follows:

$$Score_3 = \frac{\sum_{i=1}^{n} \frac{x_i}{d_i}}{\sum_{i=1}^{n} \frac{1}{d_i}} \tag{3}$$

In this method, the impact of each component is not affected by the maximum distance of the reference tree; impact is determined purely by distance. In this method the distance between the items has a great effect for small distances and, as the distance gets larger, the impact slowly falls.

In Method 4, when the evaluation item represented in the hierarchical structure and the chapter of the component are the same, the degree of impact is reduced; when the represented chapter in the reference structure is different, sudden reduction in the degree of impact occurs in accordance with the distance. In addition, when a similar

concept is referred to in the reference structure, the calculated degree of impact is relatively high.

5 CALCULATION OF SIMILARITY

5.1 Similarity calculation

In studies of the classification of documents many methods of calculating the similarity have been proposed. In this paper, we adopt the most commonly used technique as our similarity calculation method. The general procedure for calculating the similarity is shown in Fig.3.

First of all, when calculating similarity, the text information in each document is to be determined ((1) in Fig.3). Then, by morphological analysis, this text information is resolved into morphemes and extracted ((2) in Fig.3). These are the index terms (items representing the contents of the document) [11]. One morphological analysis program is "ChaSen" [12] developed by the Nara Institute of Science and Technology. Then, the words that become dissonant are removed as unnecessary words. ((3) in Fig.3). In addition, the extracted words are weighted ((4) in Fig.3). For the weighting method, index word frequency Term Frequency (TF) and Inverse Document Frequency (IDF), or a combination of these, TFIDF, are often used [11]. Finally, the similarity between texts, which are converted to vectors or matrices by weighting, is calculated ((5) in Fig.3).

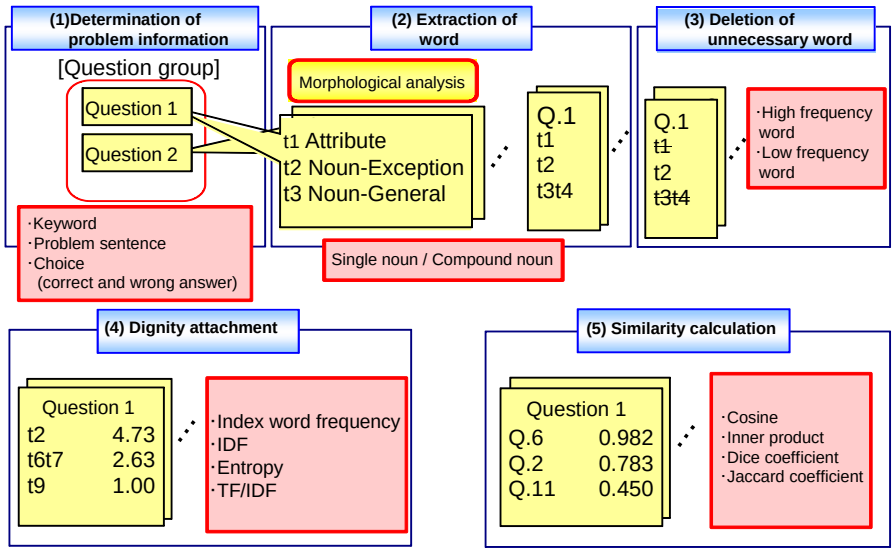

Figure 3: General procedure for calculating similarity

Table 1: Evaluation values for all methods

Category	standard value	evaluation value1	evaluation value 2	evaluation value 3	evaluation value 4
4. Information security management system	20%	11.32%	11.92%	10.45%	13.98%
5. Management responsibility	50%	13.24%	10.79%	13.36%	18.06%
6. Internal ISMS audits	0%	13.24%	10.34%	10.14%	4.91%
7. Management review of the ISMS	0%	0.00%	0.00%	0.00%	0.00%
8. ISMS improvement	0%	13.24%	8.78%	8.59%	1.84%

Table 2: Differences for all proposed types

Category	standard value	difference 1	difference 2	difference 3	difference 4
4. Information security management system	20%	−8.68%	−8.08%	−9.55%	−6.02%
5. Management responsibility	50%	−36.76%	−39.21%	−36.64%	−31.94%
6. Internal ISMS audits	0%	13.24%	10.34%	10.14%	4.91%
7. Management Review of ISMS	0%	0.00%	0.00%	0.00%	0.00%
8. ISMS improvement	0%	13.24%	8.78%	8.59%	1.84%

5.2 Application example

When experiments are carried out using a different standard, the data on the countermeasure's status in the already assessed standard are assumed.

The following application example can be considered. If the standard is updated, it is possible to locate the items of the revised chapter or the items moved to a newly created chapter. Suppose global criteria of an international standard are taken as the base. While creating the local criteria of the internal standard, the platform verifies the extent to which the underlying contents of the standard can be reflected, as well as whether any loophole has occurred. If an internal standard is provided and the aim is to obtain security attestation, the platform can be used to check how close the current internal criteria are close to the target criteria for attestation.

6 EXPERIMENTS BASED ON EACH FUNCTION

6.1 Experiment 1: Evaluation value calculation

We compared the evaluation value using the security evaluation method, which adds a weight factor to each item paying attention to the number of items and distance of a reference tree using Methods 1–4. In addition, we found that there was an impact on the results for items, when items in other chapters were being referred to within a reference tree.

6.1.1. Outline of experiment

First, we asked an evaluator who has expert security knowledge to evaluate the security of an organization, and we summarized the results in a table for every category. Next, we used Methods 1, 2, and 3, evaluated for the same security countermeasures, and compared these evaluation values with the evaluator's assessment. We also investigated whether an improvement in a value could be obtained by using Method 4, based on the knowledge acquired from the experiment.

6.1.2. Experimental results

1) Calculation of evaluation values using Methods 1, 2, 3, and 4

We input the security countermeasures into the platform, and calculated the evaluation value by each method. These values are called evaluation values 1, 2, 3, and 4. The results are shown in

Table 1.

2) Comparison of evaluation values

We compared standard values with evaluation values 1, 2, 3, and 4, and we investigated which method gives a value closest to the standard value in each management field. The differences from the standard value for each evaluation value and each category are shown in Table 2 as differences 1, 2, 3, and 4. Since a lower absolute value of difference indicates a result that is closer to a standard value Method 2, out of methods 1–3, is the most effective in Category 4. This means that countermeasures for the items of the category are in place. On the other hand, Method 3 is the most effective in Categories 5, 6 and 8, which means that reference items instead of the item of the category are being addressed. It was never the case that Method 1, 2 or 3 was the most effective in all the categories. Therefore, we used Method 4 as an impact calculation method in the form where the features of each method were harnessed. This produced an improvement in all categories.

6.2 Experiment 2: Sample presentation

6.2.1. Outline of experiment

We investigated the correspondence of the countermeasures to items of standards by showing that sample data could be generated by administrators who do not have in-depth knowledge of security attestation. This experiment was executed in the form of role play. The sample data were generated by an author who had experience in general security operations and knowledge of security standards. Countermeasure data were generated by a graduate student in our laboratory who has general knowledge about security but does not have in-depth knowledge of security standards.

6.2.2. Experimental results

1) Analysis of the countermeasures by the administrator

First, we asked an administrator to manually distinguish items of standards corresponding to the countermeasures. Since the administrator's knowledge of security standards was not sufficient, he chose items focusing on his notion of a countermeasure. Therefore, the selected results have many effective items for every countermeasure.

Then, the same task was undertaken while viewing reference-related information in a reference tree. Items with low relevance compared with the main items in each management measure were rejected. The same task was undertaken once again while viewing the sample data. The sample data were displayed in two forms, in which the data generated when extracting a countermeasure and the data generated by the administrator were distinguishable. A further reduction in the number of items judged corresponding to the countermeasures was obtained.

2) Interview of the administrator

We interviewed the administrator concerning his changing selection criteria and the results. He was able to determine the relationship among items by using the platform and selected items with confidence after presentation of the sample data. In addition, he said that he left the data that were not in samples with confidence in his judgment in practical jobs.

Table 3: Reproduced rates and assurance of items with a relation

	Number of pertinent items	Number of extraction items	OK	FN	FP	NG	Reproduced rates	Assurance
Top category	10	8	8	2	0	0	80.00%	100.00%
Middle category	31	28	25	5	2	1	80.65%	89.29%
Bottom category	116	97	95	19	0	2	81.90%	97.94%

6.3 Experiment 3: Data conversion

6.3.1. Outline of experiment

First, we compared countermeasures from two viewpoints: "the ISO/IEC 27001 Annex A" and "an ISMS attestation standard Ver.2.0 attachment". Next, we checked the results by carrying out data conversion from each dataset. We asked a graduate student who is an administrator of our laboratory to participate in an experiment using the countermeasures adopted in our laboratory.

6.3.2. Experimental results

We used about 20 countermeasures. The number of different items with a correspondence was a little more than 120. We could obtain all patterns, including opposite selection and one side selection. By analyzing the contents of items that showed a difference, we could classify the differences into the following six patterns.

i. The contents of the item were specified in detail.
ii. The contents of the item became ambiguous.
iii. If the contents of an item at a higher level to an item differ; those items to which it points also differ.
iv. The contents are expressed differently; the meaning does not change.
v. The same contents are viewed from another aspect.
vi. An item does not belong to the same category in both standards.

6.4 Experiment 4: Pertinent information extraction by similarity

6.4.1. Outline of experiment

We calculated the similarity between two standards, the ISO/IEC 27001 Annex A (hereinafter Standard A) of the international standard and an ISMS attestation standard Ver.2.0 attachment "detailed management measure" (hereinafter Standard B) which is part of a Japanese standard. The pertinent information in the two standards is already specified. We defined items that have the maximum calculated similarity between the two standards as "items with a relation" and we checked how many specified relations were reproduced. We classified items that were not reproduced into three categories: False Negative (FN), which means they were not extracted although there is a relation; False Positive (FP), which means they were extracted although there was no relation; and NG, which indicates that the wrong item was extracted.

6.4.2. Experimental results

A comparison of the pertinent information in Standards A and B the items extracted as items with a relation is shown in Table 3. The reproduced rates exceed 80% in the top, middle and bottom categories. Each assurance has a value exceeding 89%.

We investigated the 31 errors (26 FN, 2 FP and 3 NG) to determine the cause. We found that most of the combinations that cause errors have low similarity. In the top category, which contained few technical terms, if a more suitable judgment could be made, we could transfer one of FNs to the correct combination. Also, if similar words could be correctly distinguished between items, we could also transfer the other FN to the correct combination. Each of three combinations detected as FP and NG in the middle category had similarities less than 0.5. Moreover, the NG item was extracted using only an item name, so the similarity was 1, and full match was carried out. However similarity was decreased by combining the name with a portion of the detailed description. For the FNs there were also cases which showed coincidences or high similarity of item names. Other causes of FNs were a low maximum similarity viewing from both standards A and B, or a maximum similarity viewing from one side whereas the similarity is the second or third value from the other side and could not be detected because of its small margin. In the bottom category, FPs did not appear. The two combinations in the NG category showed the maximum similarity seen from one side, and had second or third similarity values seen from the other side. We could classify most of the 19 FNs into the same two cases as for the middle category

The following knowledge was acquired from these analytical results.

i. An item which has a maximum similarity less than 0.5 does not have a related item in many cases.
ii. When a description is divided into an item name and detailed description, the similarity of the item name becomes more important.
iii. Related items can be detected in many cases if they include an item with higher similarity, even when the

maximum similarity from both sides indicates that there is no related item.

6.5 Discussion

6.5.1. Experiment 1

Through these experiments based on the thinking of an evaluator, we found an influence on achievement level in the management category from items at a large distance in the reference tree of the platform. These items make reference to items outside of the management category. In addition, we found that using reference trees is an effective way to avoid human errors, such as overlooking the influence of items referring to other categories.

Moreover, the evaluation value has been improved in Category 5. "Management responsibility" by changing a method to reflect comments from an interview. However, the difference between the participant's evaluation value and the standard result is still large. This may be because possibilities are added as evaluation criteria, or because contents other than actual evaluation criteria may be reflected in a result.

6.5.2. Experiment 2

There was a tendency for a participant with insufficient professional knowledge to select more items for countermeasures. Through an interview we found out that relationships between standards were difficult to discern for the administrator who had insufficient knowledge. We also found out that it is effective to express relationships visually using reference trees and that the presentation of sample data was useful.

6.5.3. Experiment 3

From items i, ii, and iii in Section 6.3.2, since changes may come out in countermeasures by expressional range, we recognize that it is not appropriate to simply change data. From items iv, v, and vi, we see that errors can be avoided by showing the sample data.

6.5.4. Experiment 4

We confirmed that high reproducibility can be obtained by extracting items that have relations based on text similarity. We found in particular that the assurance of the items extracted was very high. Some of the causes of errors were due to improper range division of words at the time of the analysis of wording. In addition, different words with the same meaning cannot be automatically judged because technical terms are used and the similarity is low. In spite of using a simple similarity calculation, a high reproduction rate and high assurance were obtained. So it appears that it is effective to use the technique of extracting related items to determine the similarity between standards by similarity calculation methods currently used in the field of natural language processing. Moreover, it is expected that still higher reproduction rates and assurance can be obtained by creating pertinent information using a more sophisticated technique.

Once the data for sample presentation are generated, it seems to be important to reduce FP and NG items even if FN increases. This is because we assume users who do not have much specialized knowledge. For example, the following techniques may improve the results. When the standard document is divided into item names and detailed descriptions, importance should be placed on the item name instead of employing the weighting used in our experiments. Since a standard has a hierarchical structure, the similarity and detection of relations of items in higher categories should also be taken into consideration.

6.5.5. All experiments

From experiments 1, 2, and 3 we found out that platform is effective in preventing human error. The errors that can be prevented are different in each experiment, but what is considered as the primary cause of errors depends on the complicated composition of the standards used as the base document which is one of the targets of this research. In this approach, visual correspondence was provided by using reference trees, and contributed to problem solving.

In particular, visual support was provided by reference trees in experiments 1 and 2, and this contributed to the prevention of errors. In experiments 2 and 3, visual support was provided by the presentation of the sample data, which also contributed to the prevention of human error.

Moreover, in experiment 4, by using the technique of text similarity calculation, pertinent information was extracted from the standards, even where relations between the standards are not indicated. We confirmed that pertinent information can be generated from various standards, such as a global standard and a local standard.

These experiments are highly flexible and their application is not limited to security-specific standards. However, since experiment 2 is designed for choosing the relation between security countermeasures and a standard, the security viewpoint is strongly reflected here.

7 FUTURE WORK

The sample presentation function has basic issues, such as determining a sample collecting rule and reliability. Currently, we are considering solutions based on practical use rather than technical considerations. Regarding the sample collection rule we have proposed a rule in which the data generated from the sample data are provided as a new sample. Regarding reliability, we have proposed the following method, which uses a central server, in order to improve the reliability of the sample data. If entries corresponding to the same item about the same measures are stored more than a fixed number of times, the server will automatically judge that the sample data are reliable, and adopt them as the sample data. Otherwise, the data are checked by a human and adopted if their validity can be confirmed.

We have experimented with using the phases of gap analysis and present data analysis. However, there are many phases in which security evaluation can be carried out. Some

of examples are the phase in which the detailed risk analysis is conducted, and the phase in which attestation acquisition has already finished and the PDCA cycle corresponding to the phase that carries out security evaluation has already been employed. Therefore, we will also conduct a security evaluation experiment of an organization with other phases, and examine the validity of the platform.

We conducted an experiment using a calculation of similarity, and we used a standard that has pertinent information. Nevertheless, errors occurred. We will try to avoid these errors by using semantic similarity and raising text analysis accuracy. For example, we could use the structural information (e.g., about hierarchical structure) and reference information of a standard. We are planning experiments in which we will calculate the similarity by assuming that the item name is more important, if the standard item consists of a name and detailed description.

8 CONCLUSION

In this paper, we verified, based on the experimental results, not the validity of an individual function but the validity of the whole platform. We found that the platform is effective for such problems as oversight and insufficient knowledge by using visual support that presents reference trees or samples.

In this platform, we confirmed that potential influence is expressed using reference-related information in cases where influence may be overlooked even if the evaluator has expert knowledge. In addition, we recognized that the provision of visual information by reference trees and sample presentation was effective for oversight and avoidance of misjudgment, when knowledge was insufficient.

Furthermore, we found that each function can be utilized more effectively by interlocking two or more functions, such as sample presentation and data conversion. In this study, we expressed the status of countermeasures by the two choices "done" and "not yet" for simplicity. In addition, we expect that evaluation of potentiality can be improved ascertaining the optimal rate of "not yet" if potentiality and the state of being under way are expressed by using a third choice of "doing".

We conducted experiments using two standards where pertinent information was clearly specified. We recreated the pertinent information with a high reproduction rate and high assurance. By determining such pertinent information through a similarity calculation technique, we were able to lessen the rollback of the reappraisal carried out when the standard changes. It is expected that better results can be obtained by using a more sophisticated technique.

We will continue to examine the adaptability of our platform to various phases and we will try to improve its validity.

REFERENCES

[1] JIPDEC, “The international trend of ISMS, and the actual condition of a measure <2004 edition>,” (2005).

[2] Information Management Systems Promotion Center (IMSPC), “The number transition of attestation acquisition organizations,” The attestation acquisition organization of a certificate authority exception and a prefecture level, http://www.isms.jipdec.jp/lst/ind/suii.html.

[3] IPA, “Security design evaluation supportive tool V03,” http://www.ipa.go.jp/security/fy13/evalu/cc_system/CCtool_V03/secevtoolv03.htm.

[4] Y. Takahashi, and Y. Teshigawara, “A Study on a Security Evaluation Platform Based on International Standards,” IPSJ Computer Security Symposium 2008 The 2nd separate volume of collected papers, pp. 815–819 (2008).

[5] Y. Takahashi, and Y. Teshigawara, “A Study on an Effectiveness of Security Evaluation Platform Based on International Standards,” IPSJ SIG Technical Report, Vol. 2009-CSEC-46, No.13, pp.1-8 (2009).

[6] Y. Takahashi, and Y. Teshigawara, “A Study of Security Evaluation Method Based on Reference Relationships among International Standards,” IPSJ SIG Technical Report, Vol. 2010-DPS-142, No. 53, pp.1-8 (2010).

[7] Y. Takahashi, and Y. Teshigawara, “A Study on Measures Presentation Function for Non-Professional Persons of Security Evaluation Method Based on Reference Relationships among International Standards,” Multimedia, Distributed, Cooperative, and Mobile Symposium.(DICOMO2011), pp. 127–134 (2011).

[8] Y. Takahashi, and Y. Teshigawara, “A Study on Data Conversion Function of Security Evaluation Method Based on Reference Relationships among International Standards,” IPSJ Computer Security Symposium 2011(CSS2011), pp. 666–671 (2011).

[9] Information Management Systems Promotion Center (IMSPC), “International trend "ISO/IEC 27000 family,” http://www.isms.jipdec.or.jp/27000family_20111220.pdf .

[10] ISO/IEC 27001, “Information technology - Security techniques - Information security management system – Requirements,” (2005).

[11] T. Tokunaga, “Information retrieval and language processing,” University of Tokyo Press (1999).

[12] Yuji Matsumoto, Akira Kitauchi, Tatsuo Yamashita, Yoshitaka Hirano, Hiroshi Matsuda, and Masayuki Asahara, “Japanese Morphological Analysis System ChaSen 2.0 Users Manual,” NAIST Technical Report, NAIST-IS-TR99012, Nara Institute of Science and Technology (1999).

(Received October 20, 2012)
(Revised January 7, 2013)

Yuji Takahashi received the B.E and M.E from Faculty of Engineering, Soka University in 2001 and 2003. He is currently doing his Ph.D project at Soka University. His research interests are security management and international standard of security. He is a member of Information Processing Society of Japan (IPSJ).

Dr. Yoshimi Teshigawara is a Professor of Department of Information Systems Science, Faculty of Engineering at Soka University since 1995, He began his professional career in 1970 at NEC Corporation, engaged in the design and development of computer networks. From 1974 to 1976, Dr. Teshigawara was a Visiting Research Affiliate with ALOHA System at the University of Hawaii where he did research on packet radio and satellite networks. He served Dean of Faculty of Engineering and Dean of Graduate School of Engineering at Soka University. His current interests are network security, e-learning, and ubiquitous sensor networks. Dr. Teshigawara received his PhD from Tokyo Institute of Technology, Japan, in 1970. He is a fellow of Information Processing Society of Japan as well as Japan Operation Research Society. He is a member of IEEE and ACM.