Available online at www.sciencedirect.com
ScienceDirect
Procedia Computer Science 35 (2014) 918 – 928
Procedia Computer Science

18th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems - KES2014

An extraction method of ITSS common skill knowledge using Japanese text network analysis

Koji Tanaka[a,b]*, Kazuhiko Tsuda[b]

[a]*Hitachi Government & Public Sector Systems, Ltd., 2-4-18, Toyo, Koto-ku, Tokyo, 135-8633, Japan*
[b]*Graduate School of Business Sciences, University of Tsukuba, 3-29-1, Otsuka, Bunkyo-ku, Tokyo 112-0012, Japan*

Abstract

The importance of human resource development has been recognized very well; however, it is challenging to realize effective human resource development for many companies especially in the IT service industry. The IT Skill Standards (ITSS) provides indices that identify and systematize business capabilities required for providing IT services. Although the career path model is shown in ITSS, but common skills and knowledge between the job categories and the importance of each skill and knowledge are not shown. Therefore, we extracted common skills between the job categories and computed its score using a Japanese text network analysis.

Peer-review under responsibility of KES International.

Keywords: Human resouce development; IT Skill standard; Japanese text analysis; Graph; Centrality

1. Introduction

In software development, problems such as extension of the product development period and budget overruns occur frequently. According to studies conducted by Cusumano [1] '75 to 80% of product development projects are routinely delayed and over budget'. This is a critical issue that software developers should mitigate and resolve. For example, to avoid extension of the development period and budget overruns, many companies manage their project progress, product quality, costs and other factors in a quantitative manner.

* Corresponding author. Tel.: 81-3-3942-6918.
E-mail address: tnkouji@gssm.otsuka.tsukuba.ac.jp .

Peer-review under responsibility of KES International.
doi:10.1016/j.procs.2014.08.166

However, the use of these management techniques does not guarantee that development period extensions and budget overruns will be avoided. According to research from Nikkei Computer, 68.9% of projects failed, and the ratio of failed projects wherein certain quantitative management techniques were introduced was 54.4% [2].

One of the causes is inappropriate assignment of human resources with sufficient skills for the business capabilities that are necessary for a project. According to InformationWeek Reports::Research: 2012 Enterprise Project Management[3], in a figure illustrating reasons for not delivering expected results, 'Insufficient IT manpower/expertise resources' accounts for 48％ and 'IT resources improperly allocated (skill mismatch)' accounts for 14% of the reasons.

In such a background, demands for human resources with high specialties have been increasing. This has reinforced importance of human resources development to improve company competitiveness in a strategic and systematic way. The importance of human resources development has been recognized very well; however, it was challenging to realize effective human resources development. This was because, with some exceptions abroad, there was no practical index available for clarifying necessary skills and procurement policy despite many companies (especially in the IT service industry) needed definite indices. 'Skill Standards for IT Professionals' (ITSS)[4] was developed in order to solve these problems related to skill indices.

The ITSS provides indices that identify and systematize business capabilities required for the provision of IT services. ITSS is utilized as human resources evaluation figure in Japanese many companies.

In consideration of business needs, technological specialty and originality, customer liability and global recognition, the ITSS define 11 job categories and a typical model of the career path is shown. The company implement human resources development in referring to this typical model.

However, the skills and knowledge to be necessary for every job categories are shown in ITSS, but common skills and knowledge between job categories are not shown. Therefore, We don't understand it whether it leads to effective human resources development to let human resources learn what kind of skills and knowledge. Therefore, We don't understand it whether it leads to effective human resources development to let human resources learn what kind of skills and knowledge.

Therefore, we extracted common skills between the job categories and computed its score. Initially, we extracted the relationships between job categories and skills required for them from the ITSS using network Japanese text analysis. Furthermore, we computed its score using centrality.

2. Skill standards for IT professionals

The IPA has identified three types of skill standards for ICT professionals [5]: the ITSS for people working in the IT services industry, the Embedded Technology Skill Standards for embedded software development engineers and the Users' Information System Skill Standards (UISS) for information system users.

The ITSS provides indices that identify and systematize business capabilities required for providing IT services. The definition document of the ITSS consists of three parts: 'Part 1: Overview', 'Part 2: Career', and 'Part 3: Skill'. From the perspective of human abilities in a business context, the content in Part 2 and Part 3 is written with respect to 'business performance' and 'proficiency', respectively. The business capabilities required for each job category are defined by the 'Outline of Job Category and Key Performance Indicator' in Part 2 and the 'Skill Area and Skill Proficiency' in Part 3.

To identify and systematize business capabilities, the ITSS define job categories and specialty fields, establishes levels and determines indices as entry criteria for each level. In consideration of business needs, technological specialty and originality, customer liability and global recognition, the ITSS define 35 specialty fields in 11 job categories. These fields and job categories are further categorized into skill levels 1–7.

Many IT related companies make use of the ITSS. Most of them use it as an indicator of personnel evaluation and human resource development. Additionally, there are studies on the utilization of the ITSS for efficient

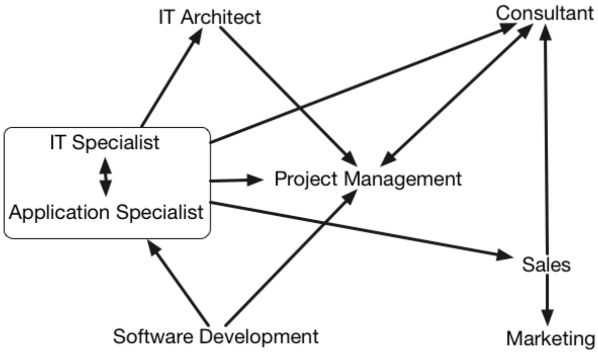

Fig. 1. ITSS Job Categories

procurement of staff when personnel change is required in a software development project [6]. This study shows that the ITSS are effective for determining if staff members have the skills required for software development.

The IPA identifies the relationships of each job categories as an example of career path models[7]. Using these career path models, the relationships of Application Specialist, IT Specialist, Software Development, IT Architect and Project Management are examples of development-related career paths. Additionally, relationships between job categories related to Application Specialist, IT Specialist and Project Management, and those related to Marketing, Sales and Consultant are shown. Fig. 1 shows a simplification of the above relationships.

In addition to the abovementioned job categories, among the 11 defined in the ITSS, three job categories, Customer Service, IT Service Management and Education, are not identified as having relationships with others. In the next section, we analyse the ITSS documents relating to the eight job categories shown in Fig. 1.

3. Creation of business capability map

3.1. Extraction of relationships and topics using compound nouns

It is easy to imagine that in documents defining skills required for an 'Application Specialist', nouns such as 'application', 'application development', 'application design', 'application architecture' and 'application security', as well as other compound nouns involving the word 'application' are used very frequently.

However, when considering the topic a document may cover in which the compound noun 'application architecture' is used, it is possible that the document may be regarding 'application' as well as 'architecture'. If the document is regarding 'application', then nouns and compound nouns involving the word 'application', as described above, will be used frequently. If the document is regarding 'architecture', compound nouns involving the word 'architecture', such as 'system architecture' and 'infrastructure architecture', will be used frequently.

Similar to the above analogy, if the compound nouns 'application design' and 'architecture design' are used in the respective documents, we can conclude that both documents also cover the topic 'design' because this word is used in both compound nouns.

Thus, it is possible to conclude that the same compound nouns appear frequently in documents that cover the same topic, and that compound nouns involving the same word appear frequently in documents that cover similar topics. Therefore, if we extract the compound nouns from documents and focus on the relationships among the compound nouns and the nouns that comprise them, we are able to understand the relationships of the documents.

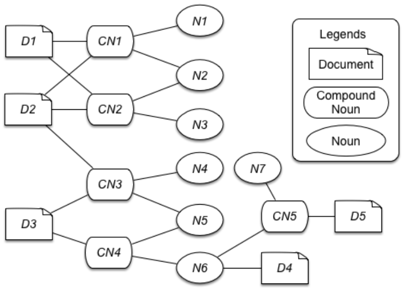

Fig. 2. Example of Created Graph

Since many compound nouns in Japanese are created by simply connecting nouns without making any changes, it is particularly effective to focus on these compound nouns. The following shows some examples of Japanese compound nouns that are not considered compound nouns in English.

Compound nouns omitting conjunction. For example, ‘Research and Develop’ in English but ‘Research Develop’ in Japanese.

Phrases constructed by an Adjective + Noun in English but expressed as Noun + Noun in Japanese. For example, ‘Environmental technology’ in English, but ‘Environment technology’ in Japanese.

The relationships among documents, compound nouns and nouns described above can be expressed in a graph, as shown in Fig. 2 below:

Documents *D1* and *D2* are connected via compound nouns *CN1* and *CN2* because both are used in documents *D1* and *D2*. Documents *D2* and *D3* are connected via the compound noun *CN3* because *CN3* is used in both documents. Documents *D3* and *D4* are connected via the noun *N6* because *N6* is used in both documents (note that *N6* is used as part of *CN4* in *D3* and is used individually as a noun in *D4*). Documents *D3* and *D5* are connected via compound nouns *CN4* and *CN5* and noun *N6* because *N6* is used in both *CN4* and *CN5*, which are used in *D4* and *D5*, respectively.

From these relationships, we can conclude that *D1* and *D2* have the most similar content, followed by *D2* and *D3*, which have less similar content. The third closest relationship is between *D4* and *D3*, followed by *D5* and *D3*.

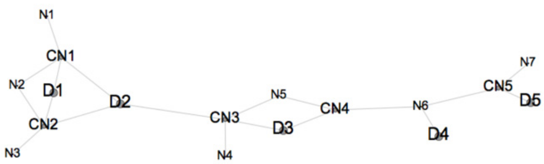

Fig. 3. ForceAtlas2 Layout

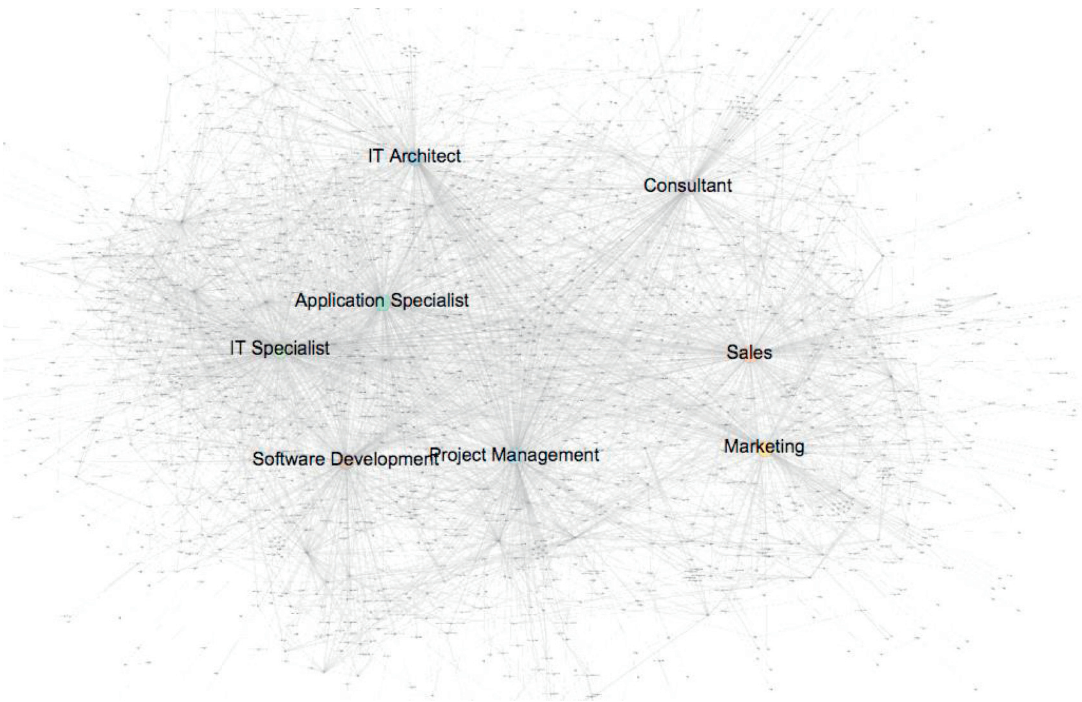

Fig. 4. Business Capability Map

The ForceAtlas2 Layout[8] can be used as a visualization technique to express the strength of relationships between each node in a graph relative to the distance between nodes. Fig.3 shows the visualized version of the graph in Fig. 2 generated using the ForceAtlas2 Layout. The distance between each document and the nouns connecting the documents can be visualised easily using this technique.

3.2. Creation of business capability map using ITSS documents

We first prepared definition documents for the eight job categories identified by the ITSS by concatenating the content of 'Outline of Job Category and Key Performance Indicator' and 'Skill Area and Skill Proficiency'.

We created graphs as shown in Section 3.1 for the definition document for each job category and generated visualizations of the graphs using ForceAtlas2 Layout.

The results are shown in Fig. 4, which we refer to as the Business Capability Map.

Additionally, to improve the visibility of the Business Capability Map, we increased the size of the labels and nodes for the definition documents of the eight job categories.

4. Extraction and scoring method of common skill and knowledge between job categories

4.1. ***Extraction of common skill and knowledge between job categories***

On the business capability map, Fig. 5 shows the example the relation between job categories and their common skill and knowledge. Fig. 5 (a) is the example when the compound noun is directly linked to the 2 job categories link directly. ITSS document defines that both "IT Architect" and "Application Specialist" require the knowledge about data modeling. Therefore, the “IT Architect” node and “Application Specialist” node are connected via the “data modeling” node. Then, the shortest path length is 2. Fig. 5 (b) is the example when the 2 job categories are directly linked to the 2 compound nouns via their common noun. ”IT Architect” shall require the knowledge of

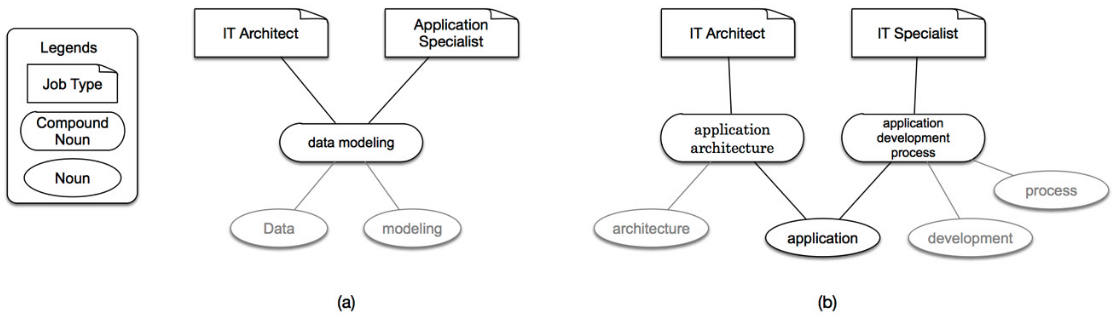

Fig. 5. Example of the Relationship of Job Categories and the Common Skills and Knowledge

application architecture and “Application Specialist” shall require the knowledge of application development process. In this case, since both the application architecture and application development process include the noun, “application,” they are connected via the application node with the path length 4. It indicates that both “IT Architect” and “Application Specialist” require the knowledge about the application and moreover it can be considered that the application architecture and application development process have the common knowledge.

As described above, in order to extract the common skill and knowledge between job categories, it shall be sufficient to list the paths between job category nodes the length of which are 4 or less and to extract the nodes between these paths.

4.2. *Scoring each skill and knowledge*

On the business capability map, each node as skill or knowledge is scored. In this case, considering that software industry is rapidly changing and the wide skill for desired human resource is required, it is assumed that the skill and knowledge applicable to more job categories would have higher score.

First in Fig. 5, the compound noun is directly linked to the 2 job categories in Fig. 5 (a), and the 2 job categories are linked to the 2 compound nouns via their common noun in Fig. 5 (b). In this case, it is considered that the compound noun directly linked to the 2 job categories in Fig. 5 (a) is more important than the compound noun linked to the 1 job category in Fig. 5 (b). Therefore, the following first rule is defined:

Rule. A : The closer a compound noun is to many job categories, the higher its score is.

Then the score *S(v)* of a node v is given as follows:

$$S(v) = \frac{1}{\sum_{j \in V} d(v,j)} \text{……Formula (1)}$$

j is a job category node, *d(v, j)* is the shortest path length from the node *v* to the job category node *j*. Then, the graph in Fig. 6 (a) shows *S(CN1)>S(CN4)>S(CN2)=S(CN3)*.

Next, the graph in Fig. 6 (b) is discussed. In Fig. 6 (b), similar to Fig.5 (b), the 2 job categories are linked to the 2 compound nouns via the common noun. CN2 is linked to the 2 nouns but different from Fig. 5 (b) CN3 is linked to the 3 nouns. Then, *S(CN3)>S(CN2)* shall be assumed because it is considered that CN3 has wider knowledge (wider use of human resource.) Therefore, the following is defined:

Rule. B : The closer a compound noun is to many nouns, the higher its score is.

Then, the part to calculate the length to the job category node in Formula (1) is expanded to calculate the length to any node *n*.

$$S(v) = \frac{1}{\sum_{n \in V} d(v,n)} \text{……Formula (2)}$$

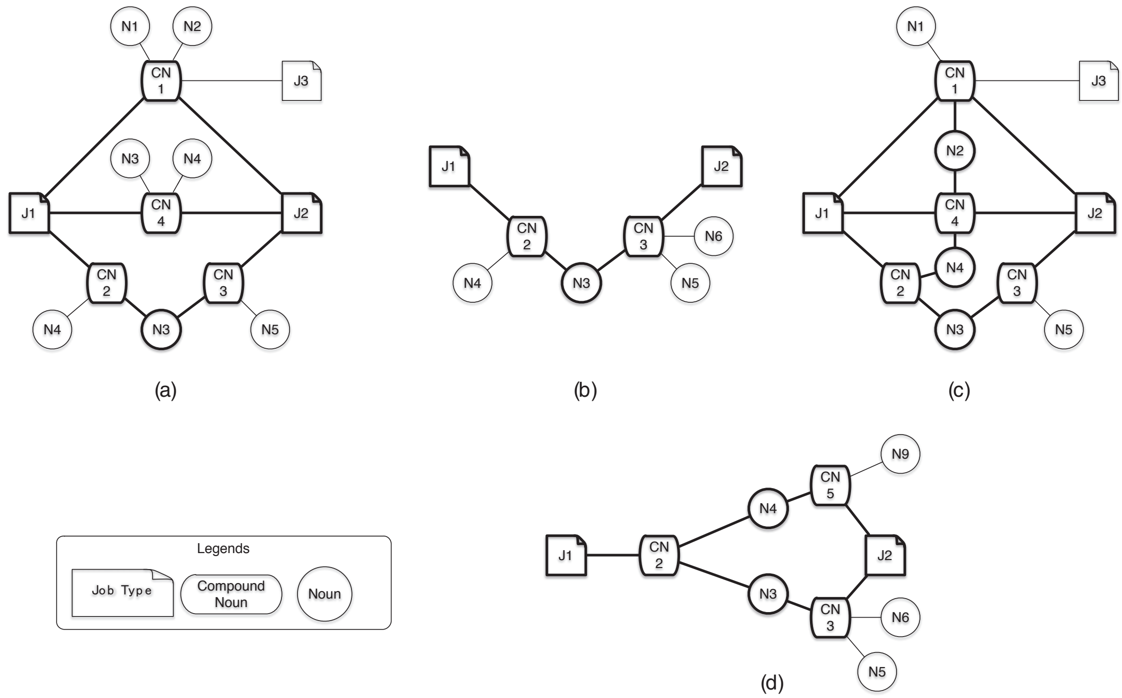

Fig. 6. Sample Graphs

The scoring of a node by Formula (2) is known as closeness centrality.

Next, the graph in Fig. 6 (c) is discussed. Every noun, N2, N3, or N4 is linked to the 2 compound nouns respectively. Here the compound nouns, CN1, CN2, CN3, and CN4, linked to them are considered, Rule. A introduces:

S(CN1)>S(CN4)>S(CN2)=S(CN3)

Then, the following is natural:

Rule. C： The more the number of the direct link of a noun to the compound noun with high score is, the higher its score is.

Then, the graph in Fig. 6 (c) shows *S(N2)>S(N4)> S(N3)*.Similarly, Fig. (d) shows *S(N3)>S(N4)* because Rule B introduces *S(CN3)>S(CN5)*.

Finally, considering CN2 and CN3 in Fig. 6 (c), it was S(CN2)=S(CN3) in Rule. A. Both CN2 and CN3 are linked to the 1 job category and the 2 nouns, but the scores of the nouns linked to *S(N4)>S(N3) >S(N5)* are different with Rule. C. Then, it is natural to consider that the score of the compound noun is higher linked to the noun with higher score. Therefore, the following is defined:

Rule. D：The more the number of the direct links of a compound noun to the noun with high score, the higher its score is.

It let the scores of CN2 and CN3 in Fig. 6 (c) be *S(CN2)>S(CN3)*.

Whether above rules were true on Formula (2) on each graph in Fig. 6 was conformed, and consequently Rule. A and Rule. B were not true. Therefore, it is determined to weight d according to the category of the linked node.

Concretely, assumed that the weight between n-th and (n+1)th node on the shortest path course between node v, n, the shortest path length would d(v, n) would be

$$d(v,n) = \sum w(p_n, p_{n-1}) \ldots\ldots \text{Formula (3)}$$

Table 1. True situation of rules with changing weight

w(s,k)	2.0	1.9	1.8	1.7	1.6	1.5	1.4	1.3	1.2	1.1	1.0	0.99	0.9	0.8	0.7	0.6	0.51	0.5	0.4	0.3	0.2	0.1
Rule.A	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	F	F	F	F	F
Rule.B	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**
Rule.C	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**	**T**
Rule.D	F	F	F	F	F	F	F	F	F	F	F	**T**	**T**	**T**	**T**	**T**	**T**	F	F	F	F	F

T : True, F : False

and the weight is changed according to the category combination of nodes p_n,p_{n+1}.

Then, assumed that a job category node is j, a compound noun node is s, and a noun node is k, it is determined that the weight between the job category node and compound noun node is fixed to w(j, s) = 1, and the weight between the compound noun node and noun node is changed.

The result to conform whether each rule was true while changing between 0.1≤ w(j,s) ≤2 on each graph in Fig. 6 is shown in Table 1.

Within 1>w(s,k)>0.5, expected result was obtained for all rules.

5. Results

On the business capability map, the common skill and knowledge for each relation of the job category shown in Fig. 1 was extracted with the method described in the previous clause. Then assumed w (s, k) = 0.75.

Consequently, it has been recognized first that there was the common skill and knowledge for all job categories. The skill and knowledge is shown in Table 2. As the common skill, the skill about the content to be managed by a team or an individual person as a project member, such as “Project Quality Management,” “Project Cost Management,” and “Project Time Management”, and the skill about the communication, “Project Communications Management,” “2-way Communication,” and “Information Transmission,” were extracted.

Next, the extraction example of the common skill and knowledge between job categories is shown. In Japan, main career paths are the job category conversion from Software Development to Application Specialist and from Application Specialist to Project Management. Therefore, the extraction result of the common skill knowledge of Software Development and Application Specialist is shown in Table 3 and the extraction result of the common skill knowledge of Application Specialist and Project Management is shown in Table 4. Moreover, IPA shows the job category conversion to IT Architect in Fig. 1 as an example of career path. Therefore, the extraction result of the common skill knowledge of Application Specialist and IT Architect is shown in Table 5.

Table 2. Common Skills and Knowledge in all Job Categories

Compound noun	Score	Compound noun	Score
Project Quality Management	0.429	Project Integration Management	0.429
Project Cost Management	0.429	Project Scope Management	0.429
Project Time Management	0.429	Project Communications Management	0.429
Project Risk Management	0.429	2-way Communication	0.428
Project Human Resource Management	0.429	Information Transmission	0.428
Project Procurement Management	0.429		

Compared with Table 3, 4, and 5, the top 5 cases of Table 3 are also included in Table 4 and the top 8 cases in Table 3 and 5 are identical. It means that if the top 5 skills shown in Table 3 are obtained at the job category conversion from Software Development to Application Specialist, then such 5 skills are directly applicable to Project Management at the job category conversion to it. Similarly, if the top 8 skills shown in Table 3 are obtained at the job category conversion Application Specialist, then such 8 skills are directly applicable to IT Architect at the job category conversion to it. Therefore, it is considered that skill and knowledge more applicable to many job categories can work out at higher score. In addition, the above result can tell that at the job category conversion from Application Specialist, it is easier for IT Architect with more common skill and knowledge than Project Management.

Table 3. Common Skills and Knowledge between Software Development and Application Specialist

No	Compound noun	Score
1	system management technologies	0.278
2	design of development environment for industrial package	0.276
3	system management methods	0.274
4	application development project	0.273
5	generic application development project	0.272
6	business application design	0.272
7	implementation design of application package	0.271
8	a person responsible for the team of the application development project	0.271
9	management strategy and systemization strategy	0.270
10	software product development planning	0.270

Table 4. Common Skills and Knowledge between Application Specialist and Project Management

No	Compound noun	Score
1	design and development management of IT solution	0.286
2	system management technologies	0.278
3	communication environment design and operation management	0.278
4	design of development environment for industrial package	0.276
5	network management technology	0.276
6	System management methods	0.274
7	design management of IT solutions	0.273
8	Information System Management	0.273
9	Application development project	0.273
10	generic application development project	0.272

Table 5. Common Skills and Knowledge between Application Specialist and IT Architect

No	Compound noun	Score
1	system management technologies	0.278
2	design of development environment for industrial package	0.276
3	system management methods	0.274
4	application development project	0.273
5	generic application development project	0.272
6	business application design	0.272
7	implementation design of application package	0.271
8	a person responsible for the team of the application development project	0.271
9	industrial package design	0.270
10	application package specific knowledge	0.270

6. Conclusion

We performed the common skill extraction and scoring between the job categories of IT Skill Standards, using Japanese text network analysis. It can be considered that this result is applicable to effective career up and human resource development.

Future work is planned that more detailed evaluation of the scoring result will performed and the scoring method will be improved. Concretely based on questionnaire data, we intend to compare the knowledge acquisition situation of skill and knowledge and scoring result for each job category and improve the scoring method to realize more practical scoring result.

References

[1] Cusumano, M. A., 2004.The business of software: What every manager, programmer, and entrepreneur must know to thrive and survive in good times and bad. Free Press.

[2] Nikkei Computer. 2009. 800 project fact-finding (in Japanese). [online] IT Pro, Nikkei Business Publications, Inc. http://itpro.nikkeibp.co.jp/article/COLUMN/20090128/323651/ (Accessed 13 July 2013)

[3] Feldman, J., 2012.InformationWeek Reports ::Research: 2012 Enterprise Project Management. [online] InformationWeek, UBM Tech. http://reports.informationweek.com/abstract/83/8656/IT-Business-Strategy/research-2012-enterprise-project-management.html?cid=pub_analyt__iwk_20120213 (Accessed 13 July 2013)

[4] IPA. 2013a. IT Human Resources Development: Common Career/Skill Framework, IPA, Japan. http://www.ipa.go.jp/english/humandev/forth.html (Accessed 13 July 2013)

[5] IPA. 2013b. IT Human Resources Development: Common Career/Skill Framework: Documents Download, IPA, Japan. http://www.ipa.go.jp/english/humandev/forth_download.html (Accessed 13 July 2013)

[6] IPA. 2013c. What is the IT skill standard? (in Japanese)., IPA, Japan. http://www.ipa.go.jp/jinzai/itss/itss12.html (Accessed 13 July 2013)

[7] Mathieu, J., and Tommaso, V. (2011) ForceAtlas2, A Graph Layout Algorithm for Handy Network Visualization. http://webatlas.fr/tempshare/ForceAtlas2_Paper.pdf (Accessed 13 July 2013)

[8] Freeman, L. C., 1977. A set of measures of centrality based on betweenness. Sociometry, pp.35–41.

[9] Brandes, U., 2001.A Faster Algorithm for Betweenness Centrality,' Journal of Mathematical Sociology, vol. 25, pp.163−177.

[10] Tanaka, K., Takahashi, M. and Tsuda, K., 2013. Comparison of Centrality Indexes in Network Japanese Text Analysis. International Journal of e-Education, e-Business, e-Management and e-Learning, Vol3(1), pp.37–42.