Analysis of Japanese IT Skill Standards Using Text Mining Techniques

Graduate School of Systems and Information Engineering
University of Tsukuba

March 2014

Rasha Fathallah Fathy El-Agamy

Abstract

This dissertation analyzes the Japanese information technology skill standards that was first announced on 2002. Generally, IT skill standards define the professional job-related knowledge, skills, and abilities required to succeed in the digital-age workplace. They can be used as a foundation tool for developing educational curriculum, profiling jobs, recruiting and evaluating employees, and designing academic and professional certification. The Japanese government established a cohesive set of overall standards for the benefit of many different organizations involved in the training of IT service professionals, by creating an openly available resource in the public domain and presenting the basic elements of an IT skill standard. Such a resource would help to raise the overall quality of IT services in Japan. The IT Skills Standards were intended from the beginning to serve as a common framework, and were compiled to serve as a public domain resource. In terms of structure and content, therefore, they have several unique features. This dissertation has initiated an analysis study to the documents of ITSS. The results of the analysis process, according to experts in software engineering, are considered to be new discoveries and very helpful for the human resources, employers, and the developers of the ITSS. This dissertation is divided into 8 broad sections including introduction and a conclusion.

The first chapter is an introduction. This chapter describes the background and the purpose of this research. The **second chapter** is a survey. This chapter describes the fundamental concepts associated with The Information Technology Skill standards, human resources and some text mining techniques that we have used in our research. It also presents some examples of the skill standards in different regions of the world as Europe, UK, USA, and Japan. The European framework is called E-cf and it has been developed by a large number of European ICT and HR experts in the context of the CEN Workshop on ICT Skills, to be a tool to describe the skills of professionals in ICT-related roles. This chapters introduces a breif explanation of the structure of the e-cf. The UK's framework is called skill standards for information age(SFIA). SFIA is a model for describing and managing competencies for ICT professionals for the 21st century, and is intended to help match the skills of the workforce to the needs of the business. It maps out the range of skills as a two-dimensional table, by tagging each skill with a category and responsibility level. This chapters describes the SFIA's concept of skills' categories and responsibility levels. In addition, the USA's skill standards and the Japanese information technology skill standards(ITSS) will be

described in this chapter. Furthermore, the outline of precedence research is shown.

The third chapter explains the concepts of the Japanese Information Technology Skill Standards. This chapters describes structures, term definitions, and meaning of ITSS. According to ITSS, problems to be solved are indicated as "tasks". Contents of these tasks vary among jobs, ITSS defines task contents of each job as job categories. So, the structure of ITSS, basically, has two major parts: careers and skills. A career framework is s chart that represents indices of outcomes required for business. A skill dictionary is another chart that lists skills necessary to achieve outcome. When an individual performs tasks by utilizing his skills, it yields business outcomes. This chapter explains the steps that any company must apply to adopt ITSS. Furthermore, this chapter introduces a comparison between ITSS's framework and another two frameworks: E-competences and SFIA.

The fourth chapter analyzes the documents of the eleven job categories of ITSS. The aim of this chapter is to analyze the activity area of each job category that was determined by ITSS using text mining techniques. We used some of these techniques to measure the similarity function between the different job categories. Moreover, we proposed a method to sort the required training courses list for each job category that were suggested by IPA. In addition, this chapter presents a method that defines the required keywords to move from one job category to another.

The fifth chapter analyzes the documents of the thirty five special fields of ITSS. It introduces a process flow of computing the similarity between special fields. We used the TF-IDF weighting scheme to compute the weight of the extracted keywords. The term-frequency (tf) is a document specific local measure and it is not the actual term frequency but it is, the actual term frequency divided by the summation of the frequencies of all keywords in that document. This frequency measure is used as the normalization factor because longer documents tend to have more terms and higher term frequencies and also this frequency measure, measures how much that term is represented inside the document. The idf part of the weighting scheme is an Inverse Document Frequency measure. The main heuristic behind the idf measure, is that a term that occurs infrequently is good for discriminating between documents. Meanwhile, this chapters introduces a proposed method that extracts the required keywords to move between special fields. The obtained results were discussed by software experts and this chapters presents the opinions and comments of the experts.

The sixth chapter a questionnaire about ITSS. This chapters explains a questionnaire that was applied to a Japanese company. The purpose of that questionnaire is to understand the kills of the employees of that company. We used this questionnaire to evaluate the obtained results in chapter five by computing the correlation coefficient. The results of correlation coefficient have approved our results and suggested some improvements of some of ITSS's documents as project manager documents.

The **seventh chapter** analyzes some career paths that were developed by IPA. To can analyze the career paths we considered each level as a document. But, IPA have not published separate documents for levels. So, we have constructed a document for each level by collecting the sections related to every level in all the documents. Furthermore, this chapter suggests the best career path for the IT engineer who would like to be a high level project manager or consultant. It also presents a proposed career path of IT-architect . **The eighth chapter** is a conclusion. This chapter presents a summary of all the findings of that dissertation.

Acknowledgement

First and foremost, I thank Allah (subhanaho wa taala) for endowing me with health, patience, and knowledge to complete this work. praise to Allah, the Almighty, on whom ultimately we depend for sustenance and guidance.

According to my belief, the one who does not thank the people does not thank his god. So, I acknowledge, with deep gratitude and appreciation, the inspiration, encouragement, valuable time and guidance given to me by my supervisor, Prof. Kazuhiko Tsuda. His kind words always gave me the necessary encouragement to proceed on the right way. I earnestly thank him for his excellent comments and guidelines during my research period. The immense trust he imposed in my abilities was always a great source of motivation. I earnestly thank him for his contribution. I am very happy that I am his first international student. All credit and recognition goes to him. My step in my endeavor was very small, but his contribution for this research was great.

I would like to extend my appreciation especially to prof. Aoe Jun-Ichi for his help to me since I came to Japan. I extend my sincere thanks to him for his assistance in the area of research and social, as well as his dedication and enthusiasm for scientific research is unprecedented, and his vast knowledge. Sincere gratitude is expressed to Prof. El- sayed Atlam, for his professional guidance and continuous encouragement throughout my thesis work. I sincerely thank him for his support in both research work and other aspects. My special thanks to him for his great efforts since I came to Japan.. Allah only can reward you, thank you. I want to express my very special thanks and gratitude to Kasai san, who help me all the time. She was very prompt whenever I needed some assistance from her. Whenever I sent her an E-mail, it never happened that she did not reply me. I wish her happy life full of success.

I gratefully acknowledge all the members of the department of risk engineering from the head of the department, Prof. Okamoto, the professors, the workers, and the students. I also wish to thank Dr. Shah Md. Faiz who was a Ph.D. student in risk engineering department and already graduated last year. He helped me a lot in many situations. He was as my tutor but in fact he was not my tutor. He answered all my questions and all the time he was saying, do not hesitate to ask or request anything. I have to acknowledge that I've learned from him in my early PhD days that I have to form my own pressure on myself to can complete my PhD. And I did, thanks for his inspiration. May allah reward him and his family. I wish him all success in his new position. My thanks to Ms. Yuran sai, who was my tutor in the first year of the doctor course. She continued to advise and help me till she left Japan. I wish her all success in her career.

My sincere thanks also goes to all my professors and friends in Mathematics department , Faculty of science, Tanta university for their continuous support. This dissertation would not have been possible without the financial support I received from the Egyptian government. My gratitude to, Egypt for providing me a research scholarship to pursue this degree, all the members of the Egyptian culture office and to the Egyptian embassy in Tokyo for their efforts.

Thank you my lovely and kind husband Mohamed Al-khodary Shenashen you never give up strengthen my well with all love, patient and sacrifice. He always encourage me to avoid frustration and mental stress during my critical busy time. Special thanks for his continuous hard work to maintain daily life in Tsukuba and to raise our two daughters: Razan and Renad. Mohamed, Razan and Renad have always stood by me and dealt with all of my absence from some family occasion.

I thank my wonderful parents for their warm empathy, their loving, support and care since I came to this life. May Allah reward them for all good things they are doing for me. My special gratitude is due to my lovely brothers; Mohamed, Ahmed, and Mahmoud. Special thanks to my extended family, mother in law for her continuous pray for me, and ask Allah's mercy and forgiveness to my father in law. I wish good health and long life for her, Dr. Khaled , Ibrahim, Sherifa, Toha and Huda.

I have to mention all those close friends, who have been there always when I needed them , or those who interacted with me almost on daily life, I mean: Mervat El-zawy, Rehab Abd El-gawad, Shaymaa Abd El -latif, , Jihan El hatir, Marwa El- marhoumy, Dina said, Sahar Al-Nadi and Ahmed Abd El- hamid.

Finally, my thanks go to all the people who have supported me to complete the research work directly or indirectly. My deepest gratitude goes to all those whom the space is not enough to mention.

Rasha El-Agamy

Dedication

To the faithful thousands that have been killed, injured or arrested defending their choice and freedom

To those who defend freedom everywhere

To those who are working to make the world more peaceful

Table of Contents

List of Tables   Pages

List of Figures

CHAPTER-1

Introduction

The information technology industry (IT) is a highly specialized area dominated by skilled professionals. With the rapid pace of developments in the business landscape, the industry has, more than ever, recognized the importance of effective human resource (HR) management. Developing a talented pool of IT professionals is crucial in sustaining business competitiveness. This is the premise behind the Japanese Information Technology Skill Standards (ITSS) developed by the Information Technology Promotion Agency (IPA) of Japan. ITSS can be used to evaluate existing skills versus market requirements. Establishing a set of key performance indicators (KPIs) can also help organizations to design their rewards and recognition programs, succession planning, and promotion. IPA defines ITSS as “a set of systematic indices that clarify and systematize the skills needed for people working in the IT services industry.” It is a tool for HR development that organizes the IT services industry into 11 job categories and 35 specialty fields. The principal targets of ITSS are IT engineers with high level of business capability in organizations.

The purpose of this dissertation is to analyze the Japanese skill standards for IT professionals (ITSS) to can extract important information that helps IT engineers to build their future and helps ITSS's developers to improve the contents of ITSS's documents. This dissertation has initiated an analysis process for some of ITSS documents using text mining techniques. This analysis process includes extraction of important keywords, using these important key words to sort the required courses for IT engineers, determining the required keywords to move between different job categories or special fields, suggesting career paths for engineers, and so on. This thesis is divided into **8** chapters including introduction and conclusion;

Chapter 1 is an introduction. It shows the definitions of important concepts and it outlines the purpose of this research.

Chapter 2 is a survey to describe the fundamental concepts associated with The Information Technology Skill standards, human resources and some text mining techniques.

This chapter includes a basic research review about the need of skill standards to deal with the current and future IT industry. Moreover, this chapters shows that, with the rapid pace of developments in the business landscape, the industry has recognized the importance of effective human resources and skill standards.

Chapter 3 explains the concepts of Information Technology Skill Standards and explains the concepts of the Japanese ITSS. In this chapter, the fundamental structure of ITSS, the career framework, and declares the people and organizations that can benefit from ITSS are described. ITSS is based mainly on two concepts: career and skills. This chapters shows the different components of each concept and the relation between these components. As the principal target of ITSS is the high level IT engineer, the process of human resource investment has been explained. To utilize ITSS in any company, there are some specific steps must be applied and must undergo ITSS adoption process. This process is composing of specific steps that must be applied to the company to can utilize ITSS in that company. These steps have been explained in this chapter.

Chapter 4 analyzes the documents of the eleven job categories of ITSS and to analyze the activity area of each job category that was determined by ITSS using text mining techniques. We used some of that techniques to measure the similarity function between the different job categories. Moreover, we proposed a method to sort the required training courses list, that were suggested by IPA, for each job category. These training subjects individuals ought to complete the aim at acquiring knowledge of a job category as explained in chapter 2. In order to improve accuracy, it is necessary to define the weight of the extracted keywords. The higher weighted keywords indicate the significance and priority of learning the associated skills. Keywords with a high weight are important for identifying the required skill for different job categories. So, it is necessary to learn the skills related to the high weight keywords. Using these high weight keywords, the required learning courses for ITSS job categories were sorted. This sorting allows IT engineers to know what are the most important courses have to learn first, so they can concentrate on studying the most needed courses for their career instead of wasting time in studying unimportant courses for them. However, the clarification of the required skills for each job category is important for IT human resources, therefore the clarification of the skills required to move between different job categories is needed. The authors have suggested a method that derives the required keywords to move between different job categories.

Chapter 5 analyzes the documents of the thirty five special fields of ITSS and text mining techniques were used to analyze these ITSS documents. In other words, a clustering engine will be built to classify documents defined by the ITSS. In order to improve accuracy, it is necessary to define the weight of the extracted keywords. We used the TF-IDF weighting scheme to compute the weight of the extracted keywords. The term-frequency (tf) is a document specific local measure and it is not the actual term frequency but it is, the actual term frequency divided by the summation of the frequencies of all keywords in that document. This frequency measure is used as the normalization factor because longer documents tend to have more terms and higher term frequencies and also this frequency measure, measures how much that term is represented inside the document. The (idf) part of the weighting scheme is an Inverse Document Frequency measure. The main heuristic behind the idf measure, is that a term that occurs infrequently is good for discriminating between documents.

Chapter 6 is an application questionnaire about ITSS. This chapter can be considered as an evaluation of the results obtained in chapter 4 compared with the results of a questionnaire that was applied on a Japanese company that uses ITSS to develop its human resources. The purpose of this questionnaire is to grasp the skills of the employees with what done in chapter 5. The correlation coefficient has been computed to know the relation between the results of the questionnaire and the results in chapter 5. The correlation coefficient value between the whole two results is 0.67. . This value is good, so the relationship between the result of chapter 5 and the results of questionnaire is high. These results indicate that the results of our previous work is near to reality. Moreover, the field-field correlation coefficient has been computed. The high correlation values means that the contents of the documents of ITSS is good and represent the reality. However, the low correlation values mean that the documents of ITSS should be improved to reflect the reality. As example, correlation value for Marketing and sales is high, this indicates that the contents of them in ITSS is clear. However, project management has low values so, the content of them should improve. The obtained results can may help the Japanese ITSS team work to overcome and develop the weak points in the documents. The development can be done by cooperation between IT professionals, companies representatives, educational institute representatives and many others occupations that are interested in ITSS.

Chapter 7 analyzes some career paths with some examples. These examples of career paths were published in a document promoted by IPA. In these career paths, changes in occupation are based on the assumption that the individual involved will remain at the same skill level before and after the change in occupation. In reality, an individual will generally descend to a lower skill level following a change in occupation. Moreover, it is assumed that the person must receive the necessary education and training in order to steadfast at the same level of skill. In ever career path there are an arrow between each level and another level. We computed the value of every arrow using text mining techniques . Then, using these values we suggested the best way to be a project manager or consultant or IT architect. Finally, from the analysis process of these career path examples, a career path for an IT architect was suggested.

Chapter 8 focuses on conclusion and possible future work for this thesis.

CHAPTER-2

Survey

This chapter describes the fundamental concepts associated with The Information Technology Skill standards, human resources and some text mining techniques that we have used in our research.

2.1 TECHNOLOGY, INFORMATION AND GLOBALIZATION

Information technology (IT) is transforming the world. A major shift in the way we live, learn and work has already begun with the arrival of the information society: students use computers at home and school to do homework; governments supply information and services on line; businesses and consumers make transactions through the Internet. This phenomenon is evident not just in the advanced countries such as United States of America and Japan but is rapidly spreading across the world.

The Internet and IT have the most prominent influence on more educated, skilled and ambitious people, especially those, are regularly working with information and communication technology (ICT). Since they are also the ones that occupy important positions in organizations hierarchy and are therefore of great significance for the overall success of the company, this study will focus on the influence of the Internet on such employees. Any potential advantage of the Internet-usage, that a company can exploit to recruit, develop and retain these types of personnel, is even more important due to the fact that there is a shortage of highly profiled people in the workforce market. [Baloh, et al., 2003]

"Information technology is a medium that permits the expression of a vast array of information, ideas, concepts, and messages…" [NCR, 1999]. According to [McClure, et al., 2003], information technology consisted of five basic elements: 1) the delivered services to an institution, such as electronic mail, conferencing and payroll; 2) the hardware and software technologies that supported the delivered services; 3) the professionals that managed the technology, services and supported others' use of them; 4) the financial resources that supported the services, technology, and professionals; and, 5) the institutional culture that dictated the first four elements. understanding and applying information technology .

In the first document published in the process of developing the first version of standards, the NWCET emphasized the importance of determining a working definition of IT which taken from John Viulamil's book, *The World of Information Technology*, is "all the technologies used for creating, abstracting, visualizing, presenting, collaborating, communication, and otherwise managing the flow of information" [NWCET, et al., 1996].

SIGITE initiative has defined the Information Technology (IT), in its broadest sense encompasses all aspects of computing technology. IT, as an academic discipline, focuses on meeting the needs of users within an organizational and societal context through the **selection, creation, application, integration and administration** of computing technologies [SIGITE, 2005].

According to Baloh, et al. [Baloh, et al., 2003], American research shows, that 79% of companies from the Global 500 group (500 largest world companies by revenue) at least to a certain extent use the Internet for seeking new personnel. Also, comparison of data for years 2000 and 1998 shows that percentage of companies that use the Internet for mentioned purpose has grown considerably from 29% in 1998. According to [RIS, 1999] report, only 4% of companies have used Internet for recruiting in the same fashion. The main advantages for Internet supported recruiting are [Survey, 1998]:

- Lower costs of recruiting (savings in invitations for application, postal-costs, data-processing costs…).
- Quicker process of recruitment: period from the point when the need for a new employee is sensed until the point when he starts doing his job is, according to the research, cut for twelve days.
- Possibility to attract better and more candidates – invitation for application published on a website can also be spotted by those, who are currently not seeking new employment actively.

Baloh, et al. [Baloh, et al., 2003] have addressed that information, communication technology and the Internet have not affected only the IT professionals and those employees using it for their work on a regular basis but also the environment of the organization, organization itself and the "social universe" [Drucker, 2001]. Managers need to be aware of these changes and try to sense them in advance and adapt to them appropriately.

Cole-Avent, et al. [Cole-Avent, et al., 2008] discussed the current state of information technology use in college student affairs from information technology leaders' point of view. The findings of this research showed that over 80% of the information technology leaders reported providing technical, educational, and training support to college student affairs professionals with specific computer applications, hardware use, and action items. College student affairs professionals exhibited fluency, through their attitude, behavior/action, and knowledge. Many of them were categorized as late majority adopters waiting for the critical mass to first adopt an innovation. The consideration of technology skill standards and the benefits to the profession received an affirmative response from the participants; however, dissension was voiced through qualitative results. Finally, information technology leaders and college student affairs professionals were identified as the two groups primarily responsible for supporting educational opportunities that promote the acquisition and development of technology skills with fluency.

2.2 INFORMATION TECHNOLOGY& HUMAN RESOURCES

Recently, great hopes are placed on information technology as the sources of industry regeneration and new economic growth. This is because the roles of IT have been expanded from the tools for manufacturing cost reduction and service speedup to those for effective collaboration among enterprises and the creation of new industries. From now on, the rise or fall of an enterprise will be determined by quality of computerization investment. Therefore an urgent matter to bring up engineers who construct advanced information systems and those who utilize them.

Silva et al ., [Silva, 1997] reported that, technology, including the information revolution, and globalization continue to exert major effects on HRD. Many enterprises have claimed that benefits of technology have not matched the cost of investment, the reason for this in most cases is that technology has not been used productively or usefully. Technology is not productive, and does not add value unless there are people who can use it productively. Total factor productivity in the major economies is estimated to have declined since the mid 1970s, and growth has been explained in terms of labor and capital inputs, the contribution by technology being considered relatively minor. Silva concluded in his paper on two notes . The first one is in Asian countries, as elsewhere, priorities in HRD to be set. which includes three basic strategies: developing human resources through education and training , deploying human resources, providing the incentives to ensure that they are productively deployed. The second is that foreign investment can be used as a means of raising the stock of human capital countries which were more engrossed in political ideologies and the fear of economic imperialism.

IT assists HR professionals in the delivery of services and affects all HR practices [Hendrickson, 2003]. Each IT tool can be used by different HR functions. For example, web data bases are used for learning at work, decision making and completing works [Benson, 2002].

Nazari et al., [Kamran Nazari, 2012] mentioned that the role that information technology can play for Human Resource Management has reviewed And using it to perform better on tasks for the director of human resources. Otherwise, the performance of human resources especially in terms of achieving the goals will be weak. So for better human resource management tasks, the new environment that is influenced by IT should be studied.

Gardner, et al., [Gardner, 2003] shows that accelerated investment and innovation in information technology (IT) offers prospects for conducting business in ways that are radically different from the past. Despite the growing presence of IT within organizations, there are not a clear understanding of how IT impacts the role of professionals. This research addressed this issue by investigating how jobs in one professional occupational segment, human resources (HR) professionals are influenced by extensive use of IT within the human resource department. They examined how HR professionals handle HR information as well as the expectations placed on them resulting from an increased reliance on IT. Their findings suggested that IT enables HR professionals to be more efficiently access and disseminate information influences by what is expected of them.

Sharma, et al., [Sharma, 2012] investigated that globalised environment, technology based short product cycles, market growth has led organizations to tailor human resource management functions. Organizations have to review external and internal environment continuously with implementation change which required to be growth oriented and competitive.

Adewoye, et al., [Adewoye, 2012] has reported that the overall and generalized impact of IT on HRM in the Nigeria Banking Sector by exploring some aspects of HRM that have been affected by IT and the effect of adoption on HRM activities through primary data collected with a structured questionnaire administered to selected Banks in South-West Nigeria. It was revealed that IT has significantly increase the efficiency of HR management activities and processes through an effective and efficient employee communication and engagement while the roles and skills of HR managers has expand considerable overtime due to their adoption and continuous upgrade of knowledge in the use of IT in the discharge of their primary functions. Efficiency of HR management activities and processes was 49% before the adoption of IT in the managerial process and 85% after the adoption, employee communication and engagement was 49% before the adoption of IT infrastructures in the HR managerial process and 79% after the adoption while role and skills of HR managers was 43% before the adoption of IT infrastructures in the whole HR managerial process and 78% after the adoption .

Mishra, et al., [Mishra, 2010] investigate the extent and comparative impact of IT use on HRM functions in organizations from different sectors. This approach has also taken into account the usage pattern of different IT tools to perform different HRM functions in organizations. Based on the survey data, the results indicated that IT has significant impact on all sectors in terms of management and planning tasks, secondly, that type of IT used varies tasks of recruitment, maintenance and development functions. The findings also support the conclusion that the use of IT is pervasive in the organizations for their HRM activities. However, there is no standardization in the integration of computer software into main HRM activities which explained the gap between job requirements and the ability of employees to perform HRM tasks. Low levels of integration of software and HRM functions can be related to fear based on ignorance and low levels of knowledge and training in IT. In general, organizations do not have portals exclusively for HR functions and use different computer software for similar HRM functions. This also means that these technologies are not systematically and maturely used for HRM functions included in the analysis in Turkey. With this backdrop, HR departments should spend more attention to the educate and train employees in HRM departments in the IT area.

In recent years, Bal, et al., [Bal, 2011] show that strategic human resources management has gained more importance for the organizations in recent years because human resources are seen as the most valuable assets of the organizations for gaining competitive. Human resources departments have started to play a strategic role in the organizations and all HR functions are integrated with the mission, vision and strategies of the organizations. The new HRM perspective for the 21st century requires HRM to be strategic partners of the organization that coordinates all functions and supporting the strategies by attracting and retaining the essential qualified employees.

Antoniu [Antoniu, 2010] Restructuring of the organizations in the current crisis (marked by mergers, layoffs and restrictions on activity) and change their strategies make career planning a very important process for both employees and employers. From the company perspective, the failure to motivate the employees by planning their careers can lead to hinder the process of filling vacant posts, a decrease of the staff involvement and an inappropriate use of the money allocated to training and development programs. From the employees' perspective, the lack of career planning can lead to frustration, feelings of not being appreciated by the company and non identifying the right position leads to the need of a job change and / or the company (particularly in the current crisis). Career planning is effective when the organizations use the skills and knowledge of their employees. Also, they are motivated to achieve maximum performance and be satisfied with their work, which helps the organization to achieve its objectives. The frequent manifestation of layoffs, especially in the current crisis, requires that employees demonstrate certain skills and competencies to the actual or potential employers that prove indispensable.

HR involvement is an important aspect in all stages of IT acquisition process [MISRA, 2006]. This study discussed a model that an organization can apply to assess its internal preparedness to manage the IT acquisition process. Application of the model in the cement company revealed many important reasons behind the current status of IT. The model stressed the importance of strategic and tactical level managers to understand the processes in the pre-acquisition stage and then organize a measuring tool to monitor the acquisition process. Studying only the pre-acquisition stage is the limitation of the study and therefore, in the next stage of the research the horizon of this model and apply it for the IT acquisition stage and post acquisition stage.

2.3 BENEFITS OF INFORMATION TECHNOLOGY

Technology affects organizations and work relations in organizations by enabling to access information and join people electronically [Ulrich, 2008]. With new processes and providing some benefits HRIS changes traditional HR processes is expected that HRIS will provide functionality for realization of units' objectives and goals as follows [Hendrickson, 2003]:

- Cost decrease: Effects of IT on HR costs appear in several ways. First, IT reduces costs of processes and works. For example, transforming from traditional HR to eHR reduces costs of some HR applications, such as, postal cost, announcement cost and data processing cost [Lin, 2011; Hendrickson, 2003]. Using self service technology reduces the processing costs of HR up to 75%. E-selections and e-recruiting decrease costs of staffing and selections due to reduced employee turnover, reduced staffing costs, and increased hiring efficiency [Strohmeier, 2007]. Second, using self service HR allows employees to perform their own work themselves directly. Thus, HR professionals spend less time on routine tasks [Baloh & Trkman, 2003].
- Saving Time: IT allows HR professionals to spent less time on routine tasks [Gardner, et al., 2003] and make easier to acquire and analyze information [Bell, et al., 2006]. For example, researches show that recruiting process shortens twelve days than before [Baloh, et al., 2003].
- Increase in Efficiency: Intense use of IT aromatizes and standardizes routines. HR professionals may focus less on administrative activities and more on interpreting information. HR professionals may spend more time on other aspects of their jobs. Thus, HR professional can access more information, respond the problems in a timely major from managers and employees and evaluate the complex information more effectively [Gardner, et al., 2003]. Comparing with manual processes, reducing data

errors, simplifying and fastening processes of HR practices make HRIS more advantageous [Ulrich, 2008; Hendrickson, 2003].

- Enabling communication and collaboration: IT is a tool for effective communication and collaboration. E-mail, messaging, discussion lists, videoconferencing, virtual teams, electronic workgroups, and teleworking have changed the nature of workplace communication and collaboration. These make workplace interactions possible for employees even they are not physically present in the workplace [Benson, 2002].
- Competency Management: IT tools enable HR professionals both to reach larger candidate pool and make decision making more objective and effective to employ more relevant and competent candidates by means of decision making techniques in the selection and recruiting process. Improving and shortening the recruiting process increases competencies of incumbents and as a result quality of works. At the same time, because of distance access eHR can be used to develop human capital of the organization effectively [Lin, 2011].
- Knowledge Management: Knowledge management is a systematic process of acquiring, creating, capturing, synthesizing, learning, and using information, insights, and experiences to enhance decision making [Ardichvili, 2002]. Knowledge management system is a natural extension of HRIS and HR development activities [Hendrickson, 2003].
- Structuring Strategic HR: Strategic role of HR focuses on aligning HR activities with HR strategies [Conner, et al., 1996]. So, HR should work with managers and line managers in collaborations [ULRICH, 1998]. IT is accepted as an important impetus for strategic HR. IT builds stronger HR units and allows HR to engage in more significant strategic roles. IT solutions free HR from the burden of routine administrative tasks [Haines III, 2008]. If HR professionals rely on IT, they hold a more strategic role and they will have time to interpret information, develop strategies and think about corporal transformation [Gardner, et al., 2003].

2.4 SKILL STANDARDS

Amid change in the workforce and economy, which has affected employer/employee workplace commitment, a need has arisen for current workers and those entering the workforce to prepare themselves for career change over the duration of their careers. Equally, employers are finding that a better understanding of their workplace skill requirements helps potential employees prepare for changes in those requirements. Skill standards are one important way to assist both employers and employees in preparing for these changes.

Skill standards have proven to be efficient foundation tools for developing curriculum profiling jobs, recruiting technical staff, evaluating employees, designing professional development programs, designing academic and professional certification. Standards provide a common-language framework for educators, business and other stakeholders to develop the educational and training tools which are necessary to prepare students and incumbent workers for today's workplace challenges as well as those that lie ahead.

With differing position titles, responsibilities, and organizational structures, the leadership has a common understanding through information technology skill standards. These standards "define the professional job-related knowledge, skills, and abilities required to succeed in the digital-age workplace. They have created a common-language for the industry and educators to develop instructional and training tools to prepare students for the workplace. The benefits of skill standards are: communicating performance expectations, reforming curriculum to correspond with workplace expectations, reducing the gap between student preparation and workplace needs, and establishing criteria for assessment, certification, and compliance [Evans; NWCET, 2000]. Different from competencies, skill standards were described as representing a higher-level of knowledge, skills, abilities, and performance (NWCET).

The Northwest Center for Emerging Technologies sponsored by the National Science Foundation, developed National information technology skill standards for eight career clusters: database development and administration, digital media, enterprise systems analysis and integration, network design and administration, programming/software engineering, technical support, technical writing, and web development and administration. "For skill standards to be effective, they must reflect the consensus of the industry professionals in that career field" [NWCET, 1999]. These standards were created from the contributions of expert panels, information technology-based companies, along with the consideration of work functions, technical knowledge, and related employability skills.

The standards were separated into a three tiered system represented by a triangle form. Tier I, the base, is comprised of the necessary foundational skills, knowledge, abilities, and personal qualities which focused on mastering basic skills such as information technology use, critical thinking, and the management of time, resources, and information. Tier II encompassed proficiency in technical skills, knowledge, and abilities with Internet techniques, and understanding and troubleshooting issues with hardware and software. Tier III addressed technical skills, knowledge and abilities identified by the information technology industry and unique to organizations. These skills included understanding industry terminology, compliance and legal requirements, organizational practices and protocols for the preparation of leaders dealing with rapid change.

Cajander, et al., [Cajander, 2011] addresses the issue of developing and assessing professional skills in higher education programs. Cajander's approach includes defining and assessing these skills, in the contexts of an individual course unit and for an entire degree program. This can be particularly important in programs that aim to offer students a truly collaborative learning experience in a culturally diverse team. Reflections are presented as one example of an assessment method that fits this requirement. Building assessment based on the notion of threshold concepts is introduced in the context of an open ended group project course unit at Uppsala University.

Stracke [Stracke, 2011] summarize the potential use cases and impacts of competences and skills in the new area, often called "Digital Age" which highlights the roles and benefits of standards and metadata for HR development and points out the special support that competence models can provide for the quality development in learning, education, and training. Stracke concludes that Competence and quality development are crucial and indispensable for the long term success of learning opportunities and in particular of vocational education and training: To reach an economical benefit through competence modelling and building, standards we should offer a sustainable support.

2.4.1 What are Skill Standards?

The National Workforce Center for Emerging Technologies (NWCET), located at Bellevue Community College in Bellevue, Washington, has identified and described skill standards for eight IT career clusters in the publication Building a Foundation for Tomorrow: Skill Standards for Information Technology. The millennium edition of this work is the end result of a national review and update of the IT skill standards, with input from expert panels around the country for new and changing skills, work functions, technical knowledge, and related employability skills.

In Japan, the Information technology Promotion Agency (IPA) [*http://www.ipa.go.jp/-index-e.html*] has identified IT skill Standards(ITSS) [*http://www.ipa.go.jp/english-/humandev/forth_download.html*] as the indices that identify and systemize business capabilities required for providing IT services. Moreover, IPA stated that the purpose of ITSS is : the capital investment of human resources which is a key issue of business management, and ITSS promotes to develop high level IT professionals with social market value and low level IT professionals to be able to establish themselves in the IT markets. Furthermore, ITSS aims at the assessment of business capabilities of HR who are working as professionals engaged in the IT service industry. This approach mainly focuses on the IT service industry providing customers with IT services, but it is also applicable to IT departments in user companies as general.

2.4.2 Why are Skill Standards Important?

In today's workplaces, the only constant is change. Jobs that once were relatively simple now require high performance work processes and enhanced skills. Because skill standards reflect changing workplace realities, they become a tool for applicants and employees to access greater career opportunities. National recognition of skill standards in career fields provides a common basis for certifying achievement against those standards, thereby allowing the portability of skills across geographic areas, companies and careers. Updating skills and knowledge is now a lifelong endeavor, causing many employers and employees to spend more effort, time and money on education and training. Skill standards provide benchmarks for making education and training decisions, shaping curricula, and directing funds toward highest value education and training investments [Allen, 1999].

Successful industrialized nations that have maintained their competitiveness are characterized by a well-established skill standards system. The application of skill standards to the development of curriculum results in courses and programs whose outcomes can be assessed across a broad range of contextual technical and foundation performance criteria to the development of curriculum results in courses and programs. These results in learners are prepared to function effectively in the technology- and information-based workplace [NWCET, 2003].

2.4.3 Who should use Skill Standards?

The skill standards should be used by:

- The nation for ensuring the development of a high skills, high quality, high performance workforce, including frontline workers;
- Industries to inform training providers and prospective employees of needed skills;
- Employers to evaluate the skill levels of prospective employees and assist with the training of current employees;
- Labor organizations to enhance employment security through portable : credentials and skills;
- Workers to obtain certifications of skills, pursue career advancement, and enhance their abilities to reenter the workforce;
- Students and entry-level workers to determine needed skill levels and competencies for the workforce;
- Training providers and educators to ascertain appropriate training services;
- Government to evaluate publicly-funded training; facilitate transition to high performance work organizations; increase opportunities for minorities and women in the workforce; and facilitate linkages with other national efforts aimed at enhancing workforce skills, such as school-to-work transitions, vocational technical education, and job training programs.

2.5 COMMON SKILLS ACROSS ALL IT HUMAN RESOURCES

This section will presents some of the employer's general requirements for specific skills:

- Common Skills

  Apply a systematic, methodical approach to solving a problem, survey who else knows about the problem, Develop a rational set of possible solutions, Test the solutions cost-effectively and efficiently, Verify that the problem is truly solved, and document the solution for others

- Technical Skills

  In addition to these common elements, specific items of technical knowledge, skills, abilities, attributes, and use of tools are associated with a function or task. These specific items are represented at a high level and avoid reference to a specific vendor, version, or piece of equipment. This process allows maximum flexiblity in adapting the skill standards to local specifications while preserving the employer's general requirements for specific skills.

- Employability Skills

  Finally, there are employability skills—general requirements associated with a function or task. Input from industry clearly shows that without solid mastery of employability skills, an employee cannot succeed in the highly competitive environment of today's technology company. Employers often say that "technical skills may get you the job, but foundation skills make you a valued employee and significantly increase career advancement." For our purposes, let us identify seven types of employability skills:

  1. Communication skills. Effective information flow throughout the organization is a critical element in organizational success. Communication with team members, supervisors and subordinates, and customers and clients as well as between groups must be timely and appropriate. Some jobs rely heavily on written communication, while others depend primarily on verbal communication. Communication in a high-technology organization takes on many forms: informal or formal presentations, technical logs, complex reports, proposals, and so on. No matter the form,

communication is vital to individual and team effectiveness.

2. Organizational skills

   As employees are asked to handle more parallel tasks with an increased level of complexity, good organizational and planning skills become important. Depending on the job, the complexity of the organizational task may vary from scheduling and prioritizing multiple tasks or requests to planning and tracking complex and capital-intensive projects involving many people and teams. Regardless of the size of the project, the ability to identify and define tasks, track milestones, recognize when a project timeline is running into problems, and take appropriate action is crucial to ongoing success in a technical job.

3. Team contribution and leadership

   Most organizations are relying increasingly on teams to accomplish projects. This is particularly true in high-tech environments where the success of a project depends on the contribution of many individuals with varied expertise. The ability to work with team members with diverse backgrounds and communication styles is highly valued and rewarded in most environments. Being able to read the needs of the team as a whole and the needs of individual team members, and to adjust one's role to increase team effectiveness, is essential to the success of the team process.

4. Professionalism. Dealing with problematic employee issues, attitudes, and behaviors consumes much time in any organization, and it can be quite detrimental to overall morale. Employees with good work ethics, who show up on time, who understand and follow company procedures, and who relate to coworkers and customers with respect are usually the ones selected for a position with an increased level of responsibility and reward.

5. Critical thinking and decision making

   As an organization becomes leaner in management, the employee is expected to assume increased responsibility. An employee's ability to correctly analyze a situation, understand tradeoffs, offer good recommendations, and make the right choice is often rewarded with increased freedom to self-manage, and with the opportunity to engage in more interesting and challenging projects.

6. Customer relations

   Customers can wear many faces. An internal customer is the department down the hall or an offshore division. An external customer is a supplier, client, or end user. The ability to solicit and listen to customer feedback and to effectively address customer issues and concerns is required to qualify for certain positions, such as a technical support job. Customer interaction skills are necessary in every job, whether or not its description formally includes "customer relations."

7. Self-directed and continuous learning In the high-technology industry—especially in an information technology environment—technologies and practices change rapidly and sometimes radically. To keep up with technology change, employees must constantly engage in self-assessment against the technological landscape of skills and knowledge and then take proactive steps toward enrolling in continual training for their trade. The employer expects employees to be current in their technical skills. Most organizations provide the necessary resources for continual training. However, it is often seen as the employee's responsibility to identify personal gaps in knowledge and take actions to fill these gaps.

2.6 SKILL STANDARDS WORLDWIDE

As we mentioned, skill standards are very important for today's IT industry, so there are many kinds of skill standards all over the world. This section we will introduce some examples of these skill standards.

2.6.1 The European e-Competence Framework

The European e-Competence Framework (e-CF) provides a reference of 36 competences as required and applied at the Information and Communication Technology (ICT) workplace, using a common language for competences, skills and capability levels that can be understood across Europe. As the first sector-specific implementation of the European Qualifications Framework (EQF), the e-CF was designed and developed for application by ICT service, user and supply companies, for managers and human resource (HR) departments, for education institutions and training bodies, and other organizations in public and private sectors [Breyer, et al., 2007; eCF, 2010; Newsletter, 2005].

The framework was developed under the umbrella of the CEN ICT Skills Workshop through a process of close cooperation between ICT business and human resource (HR) experts, stakeholders and policy institutions from many different countries and at the EU level. This framework published by CEN for the first time in 2008 and followed by a further enhanced version 2.0 in 2010, the framework brings benefits to a growing community of users throughout Europe and overseas.

The European e-Competence Framework is structured from four dimensions. These dimensions reflect different levels of business and human resource planning requirements in addition to job/ work proficiency guidelines. as shown in Figure 2-1 . The four dimensions of the e-CF are specified as follows:

- Dimension 1: five e-Competence areas derived from the ICT business processes PLAN – BUILD – RUN – ENABLE – MANAGE. This area is instrumental in HR assessment and the allocation of training programmers as well as the identification of e-competences. It helps HR managers to communicate with business managers and make joint decisions, and it is a first guide to organize and list e-competences.
- Dimension 2: A set of reference e-Competences with a generic description for each competence. It is the core of the framework. 36 competences have been included in e-CF 2.0. These competences are not business sector-specific, i.e. they do not address specific applications such as banks, health, transport etc. They are general e-competences customizable and applicable to any industry or business sector. In dimension 2, e-competences refer to and represent organization needs.
- Dimension 3: For each e-competence, suitable proficiency level specifications ranging between e-Competence levels e-1 and e-5 have been constructed. They relate to EQF levels 3 to 8. This dimension involves “behaviors” and levels of autonomy, a bridge from “organizational” to “individual” competences is created. Note that organizational competences are generic and broad whilst individual competences are specific and customized.
- Dimension 4: Knowledge and skills embedded within e-Competences are listed. They are not intended to be exhaustive but are examples1 of e-competence content. These examples can be useful, in defining specific and precise outcomes to be assessed

within an organization’s competence assessment programmers. Moreover, they provide inputs for training institutions2 to help define learning outcomes and design training initiatives. Dimension 4 components refer to Dimension 2 but are not related to specific competence levels in dimension 3. However, Dimension 3 has been used to verify the applicability of knowledge and skills identified.

Dimension 1	Dimension 2	Dimension 3				
5 e-Comp. areas (A – E)	36 e-Competences identified	e-Competence proficiency levels e-1 to e-5, related to EQF levels 3-8				
		e-CF levels identified per competence				
		e-1	e-2	e-3	e-4	e-5
A. PLAN	A.1. IS and Business Strategy Alignment				■	■
	A.2. Service Level Management			■	■	
	A.3. Business Plan Development			■	■	■
	A.4. Product or Project Planning		■	■	■	
	A.5. Design Architecture			■	■	
	A.6. Application Design	■	■	■		
	A.7. Technology Watching				■	■
	A.8. Sustainable Development			■	■	
B. BUILD	B.1. Design and Development		■	■	■	■
	B.2. Systems Integration		■	■	■	
	B.3. Testing	■	■	■		
	B.4. Solution Deployment	■	■	■		
	B.5. Documentation Production	■	■	■		
C. RUN	C.1. User Support	■	■	■		
	C.2. Change Support		■	■		
	C.3. Service Delivery	■	■	■		
	C.4. Problem Management		■	■	■	
D. ENABLE	D.1. Information Security Strategy Development				■	■

Figure 2-1: E-CF Overview

The e-CF is competence-based approach and is flexible in application. The alternative, job-profile approaches, are less flexible, making local adaptation difficult. However, combining competences from different competence areas and using them as building blocks can create flexible job-profiles. This enables the derived job profiles to be easily updated by substituting or deleting competences without the need to restructure the entire profile.

2.6.2 Skills Framework for the Information Age (SFIA)

The Skills Framework for the Information Age (SFIA) is the UK government backed high-level IT skills standard. It describes the typical roles in IT and the skills needed to fulfill them. SFIA provides a solid foundation for organizations when setting standards of professionalism in IT [http://www.sfia-online.org/; Newsletter, 2005; Bailey, et al., 2011]

SFIA is a model for describing and managing competencies for ICT professionals for the 21st century, and is intended to help match the skills of the workforce to the needs of the business. This model maps out the range of skills as a two-dimensional table, by tagging each skill with a category and responsibility level. These categories are divided into six main areas: Strategy and planning; Business change; Solutions development and implementation; Service management; Procurement and management support; and Client interface. Each of these categories is then further divided into sub-categories, mapping out 86 separately identifiable skills. Each of these skills has a general description. There are seven levels of responsibility, in ascending order: Follow; Assist; Apply; Enable; Ensure and advise; Initiate and influence; and Set strategy, inspire and mobilize. Each of these responsibility levels has a generic description showing the level of autonomy, influence, complexity, and business skills required.

From a Professional Services perspective, responsibility levels 3-7 can be used as equivalent to seniority levels: Associate professional, Professional, Senior professional, Lead professional, and Principal professional. Each level of responsibility within each skill has a brief description of the typical tasks expected of someone in that role, and as one skill maps to all seven levels (Management and Operations, within the Service Provision category), these results in 263 detailed tasks defined. SFIA enables the measurement and benchmarking of an organization's collective IT skills and its processes for managing and developing. The framework provides a clear model for describing what ICT practitioners and users do. It is constructed as a two-dimensional matrix as shown in Figure 2-3.

Skills:

One axis divides the whole of IT into "skills". Skills are grouped for convenience into subcategories or "business roles". Subcategories are grouped into six categories or work areas - strategy & planning, management & administration, development and implementation, service delivery, sales & marketing, and use. The SFIA structure allows a consistent approach of IT skills across the organization and is not limited to a specialist department.

Levels:

The other axis defines the level of responsibility and accountability exercised by IT practitioners and users. Each of seven levels - from new entrant to strategist level - is defined in terms of autonomy, influence, complexity and business skills as shown in Figure 2-2.

Descriptors:

The matrix shows the complete set of skills used by ICT practitioners and users. For each skill at each level, "descriptors" provide examples of typical tasks undertaken. A typical task for systems design at level 5 is "review others' system design to ensure selection of appropriate technology, efficient use of rresources, and integration of multiple systems and technology" as in Figure 2-3.

Figure 2-2: SFIA Levels

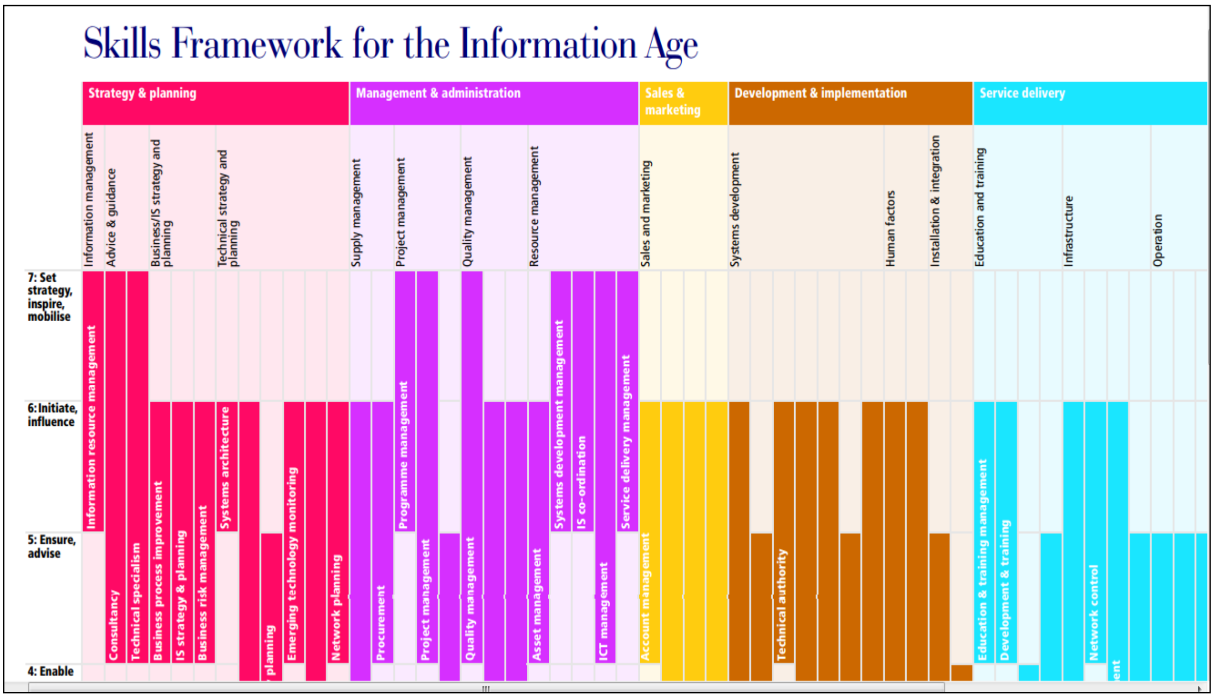

Figure 2-3: SFIA Framework

2.6.3 NWCET IT Skill Standards

In 1996, the National Workforce Center for Emerging Technologies (NWCET) and the Regional Advanced Technology Education Consortium (RATEC) identified skill standards for the original eight Information Technology (IT) career clusters in the first edition of Building a Foundation for Tomorrow [NWCET, 2003; NWCET, 1997; NWCET, 1999].

The NWCET IT Skill Standards were developed through input from hundreds of professionals in the field of IT in order to elicit a comprehensive description of all activities that occur on a regular basis in IT. These findings were verified with more than 10,000 front-line workers and first-line supervisors in the field who agreed that the framework actually describes the world of IT work. The NWCET IT skill standards are structured around four major components:

- Major *career clusters* across the entire field of IT
- Major *critical work functions* and *key activities* that characterize the work within each career cluster
- Level of performance, stated as *performance indicators*, expected from a professional in each domain of work
- *Technical knowledge* and *employability skills* necessary to carry out the work at the expected level of performance

Career Clusters

NWCET's current research identifies 8 career clusters in the world of information technology: Database Design and Administration, Digital Media, Enterprise Systems Analysis and Integration, Network Design and Administration, Programming/Software Engineering, Technical Support, Technical Writing, Web Development and Administration. These 8 career clusters within the IT professional world evolve as the technology evolves. The NWCET IT Skill Standards have changed since they were first developed in 1998, to reflect the changing nature and face of the IT work world. New technologies and trends have pushed the development of new career clusters. The second edition of the NWCET Skill Standards published in 2000, featured the addition of Web Development and Enterprise Systems

Analysis and Integration.

The third edition, being published at this time, continues to feature the same 8 job clusters, although industry feedback indicates that Web Development may melt back into the background as the web becomes so ubiquitous that it no longer makes sense to call its own career cluster. Certain clusters, like Networking and Technical Support, are revealing the impact that wireless technology having. Every cluster is being impacted by issues of security and the NWCET recently completed its first edition of IT Security Skill Standards.

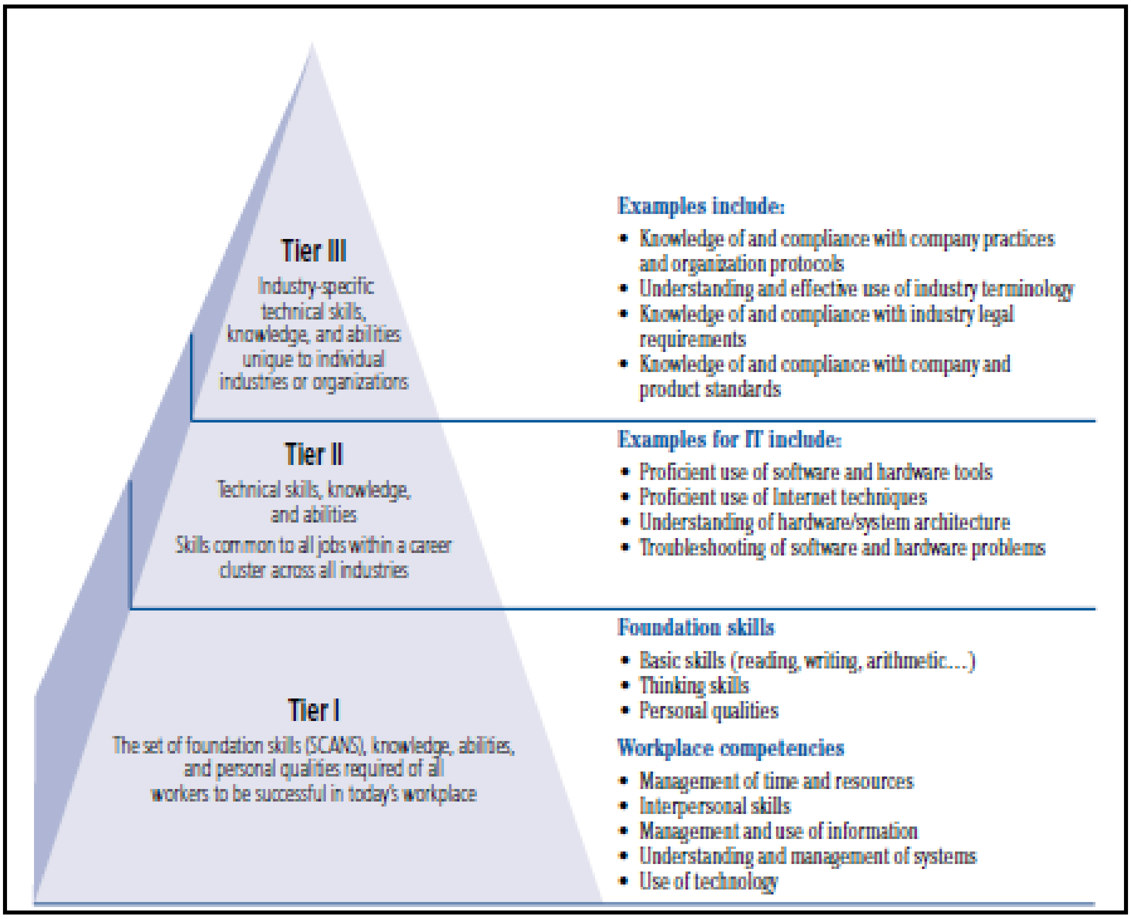

Figure 2-4: NCWET's Tires of Skill Standards

2.6.4 Japanese IT Skill Standards (ITSS)

Intensifying competition in the IT industry has prompted companies to shift their development bases overseas to avoid high personnel expenses in Japan, and is causing a so-called hollowing-out of the industry. To prevent this shift, there have been calls for urgent action to develop essential IT human resources. In response, METI established ITSS as an industry skill standard in 2002. METI subsequently transferred the control of ITSS to IPA and entrusted IPA with the administration of these standards.

Specifically, ITSS is a set of systematic indices that clarify and systemize the skills needed for people working in the IT services industry. ITSS is utilized as a tool for developing professional human resources to implement corporate strategies and organized into a career framework, as shown in Figure 2-5, ITSS classifies the information services industry into 11 job categories and 35 specialty fields. In each field, there are seven levels based on individual experience and results. One appealing feature of ITSS is that allows engineers to draw roadmaps for their own futures and career advancement(Career Path).

Approximately 90% of large enterprises and over 60% of SMEs have introduced or are considering users of ITSS. According to these figures, ITSS is being effectively utilized as an indicator for business managers and engineers to systematically consider both their own and their respective companies future. As we mentioned in chapter 1, this thesis is focusing on analysis of the documents of ITSS, so, more detailed description of ITSS's concepts and basic structure will be introduced in chapter 3.

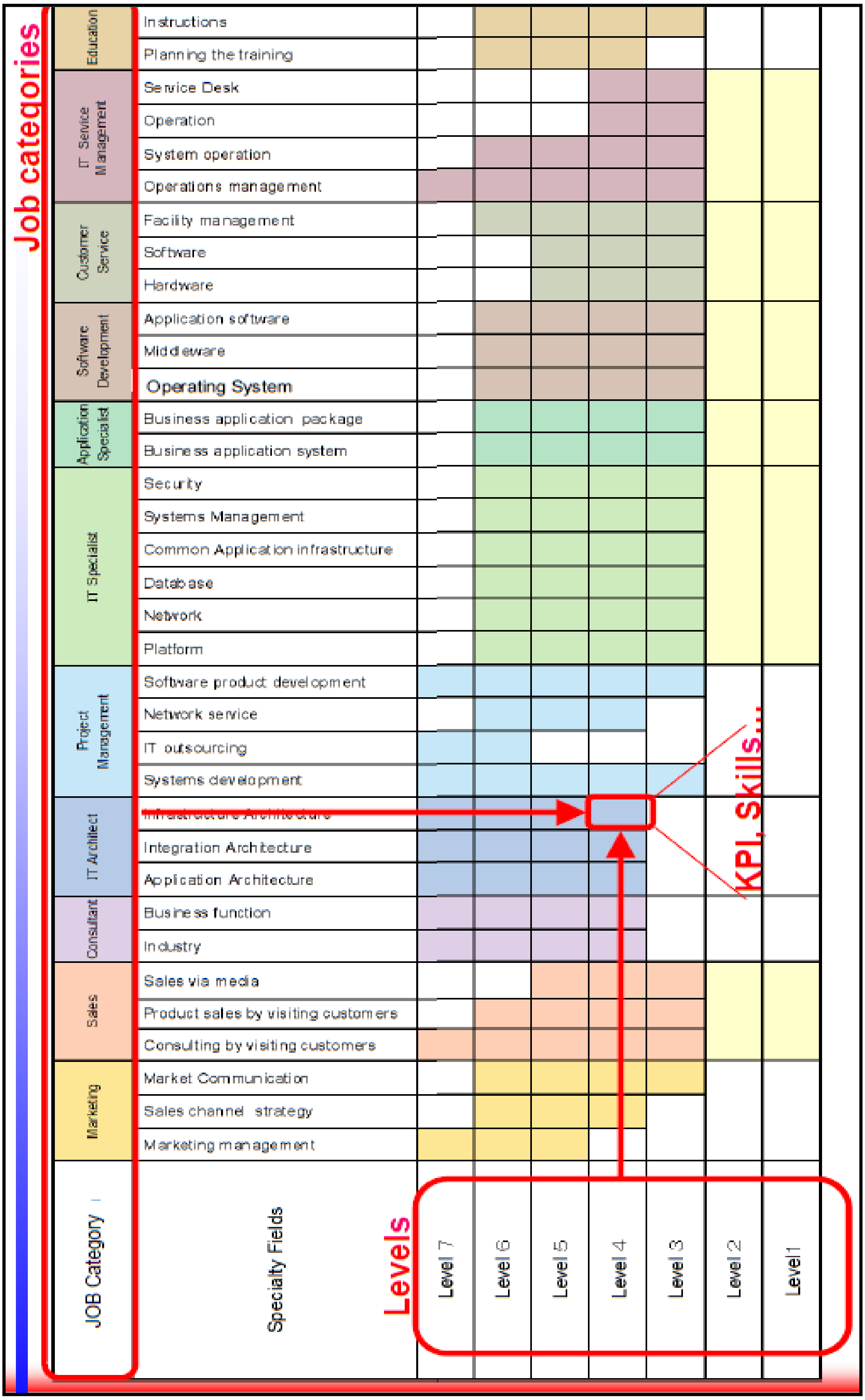

Figure 2-5: ITSS's Framework

2.7 TEXT MINING TECHNIQUES

The goal of this research is to analyze the Japanese information technology skill standards using text mining techniques as clustering, keyword extraction and so on. So, the following part of this chapter will review some of these techniques.

2.7.1 Document Clustering

Clustering is the process of organizing data objects into a set of disjoint classes called clusters. Objects that are in the same cluster are similar among themselves and dissimilar to the objects belonging to other clusters. Clustering is a crucial area of research, which finds applications in many fields including bioinformatics, pattern recognition, image processing, marketing, data mining, economics, etc. Cluster analysis is one of the primary data analysis tools in data mining [Salton, 1989; Christopher, et al., 2009; Jones, et al., 1997]. Document clustering is the task of automatically organizing text documents into meaningful clusters or group, In other words, documents in one cluster share the same topic, and the documents in different clusters represent different topics [Zhang, et al., 2009; Hamasuna, 2010; K.Sathiyakumari, et al., 2011]. Many different clustering techniques have been proposed in which few are described to produce different clusters. They are classified into following categories.

- **Hierarchical versus partitioning methods (nested and un-nested):**

Hierarchical techniques produce a nested sequence of partitions, with a single, all inclusive cluster at the top and singleton clusters of individual points at the bottom. This cluster produces a hierarchical tree structure with the leaves of tree are individual clusters of all object inputs and the cluster related to a particular node in the tree is the union of all clusters related to the child nodes of that particular node. [Reynolds, et al., 2006; Krishna, et al., 2010].

Partition Techniques create un-nested clusters where data belongs to only one subset of clusters. If K is the number of clusters, then partitional approach typically finds all K clusters at once. Partitioning method includes K-Mean algorithm where each cluster is represented by mean value of the object in the cluster and K-Mediods algorithm where each cluster is represented by one of the objects located near the center of the cluster are popular heuristic methods.

- **Density-based methods:**

The idea behind this technique is to continue growing the given cluster as long as the density (number of objects or data points) in the “neighborhood” exceeds some threshold. DBSCAN and its extension, OPTICS are typical density-based methods that grow clusters according to a density-based connectivity analysis. DENCLUE is a method that clusters objects based on the analysis of the value distributions of a density functions[Rokach, et al., 2010].

- **Grid-based methods:**

Grid-based method first covers the problem space domain with a uniform grid mesh. Statistical attributes are collected for all data objects located in each individual mesh cell and clustering, is then performed on the grid, instead of data object themselves. The main advantage of this approach is the fasting processing time. STING is a typical example of grid-based method [Ilango, et al., 2010].

- **Model-based methods:**

Model-based methods hypothesize a model for each of the clusters and find the best fit of the data to the given model. EM is a Model-based algorithm that consists of two alternating steps: the Expectation (E) step and the Maximization (M) step based on statistic modeling. COBWEB is a conceptual learning algorithm that performs probability analysis and takes concepts as a model for clusters [Feng, et al., 2007; Zhong, et al., 2003].

Traditionally, in document clustering, documents are represented by vectors of term frequencies. Many existing document clustering techniques use this simple ‘‘bag-of-words’’ model to represent a document in a collection [Voorhees, 1986]. The bag-of-words simply consists of terms that appear in a publication’s original source text. Each term is assigned by ‘‘weight of importance’’ using a weighting metric such as tf-idf weighting scheme [Salton, 1971].

2.7.2 Term Weighting Schemes

As mentioned in [Salton, et al., 1988], only certain terms extracted from a document can be used for identifying and scoring a document within the collection. Term weighting schemes can be used to identify the importance of each term with respect to a collection and assign weights to them accordingly. Document clustering uses these term weights to compare the similarity of two documents. Several term weight schemes are in use today, but none of them is specific to the clustering algorithms. In this section, we discuss the general information related to term weighting and document clustering. Traditionally, the Boolean retrieval model assigns 1 or 0 based on the presence or absence of the terms in a document. This model performs undesirably in querying for a document. Later, the Vector space model was introduced for ranked retrieval [Salton, 1989]. This approach is widely used in querying documents, clustering, classification and other information retrieval operations because it is simple and easy to understand. It uses a bag of word approach. Each document d_i in the collection D is represented as a vector of terms,

$$d_i = \{term_1, w_{1i}; term_2, w_{2i}; ...term_T, w_{Ti}\} \qquad (2\text{-}1)$$

And each term $term_t$ in a document d_i is assigned a weight w_{it} which represents its importance. The term weight determines whether the term $term_t$ will be included in the further steps or not. Several term weighting schemes have been proposed to compute the importance of a term in a document and in a collection. Norm-TF, TF-IDF, ATC, LTU, and Okapi are some of the widely used term weighting schemes [Salton, et al., 1988; Jones, et al., 1997]. One of the most commonly used term weighting schemes is TF - IDF. It measures the importance of a term using its frequency within a document and the inverse of its document frequency within the collection. Several papers have suggested modifying an existing term weighting scheme for their methods. Ayad, et al., [Ayad, et al., 2002] show a modified TF - IDF term weight to avoid single terms from getting higher weight. TF - ICF (Term Frequency Inverse Corpus Frequency) was proposed as a new term weighting scheme by Reed, et al., [Reed, et al., 2006] for clustering a dynamic data stream. It uses the existing collections to weight the terms in the data stream. Also, Zhang, et al., [Zhang, et al., 2009] used DF - ICF

Inter- and Intra-cluster components for extracting a description from a cluster.

A comparative study of various versions of TF-IDF term weighting schemes and presented a new term weighting scheme $tf.rf$ to improve the term's discriminating power has made by Lan, et al., [Lan, et al., 2005]. The controlled experimental results showed that this newly proposed $tf.rf$ scheme is significantly better than other widely used term weighting schemes. Brunzel, et al., [Brunzel, et al., 2007] introduced a term weighting scheme which improves the behavior compared to the traditional TF-IDF scheme by adding a domain relevance component that measures the degree to which a term is regarded as more relevant within a corpus compared to a reference corpus. Reed, et al., [Reed, et al., 2006] proposed a new term weighting scheme called Term Frequency - Inverse Corpus Frequency (TF-ICF). It does not require term frequency information from other documents within the document collection and thus, it enables researcher to generate the document vectors of N streaming documents in linear time. Murugesan, et al., [Murugesan, et al., 2011] proposed a new term weight scheme and investigated its use in document clustering algorithms. They introduced two new cluster components for our term weighting scheme and they have demonstrated how these cluster components in addition to the term and collection frequency components in this term weighting scheme improves the average entropy result of the K-means algorithm.

2.7.3 SIMILARITY MEASUREMENTS

As mentioned in previous sections, clustering is a useful technique that organizes a large quantity of unordered text documents into a small number of meaningful and coherent clusters, thereby providing a basis for intuitive and informative navigation and browsing mechanisms. Accurate clustering requires a precise definition of the closeness between a pair of objects, in terms of either the pairwised similarity or distance. A variety of similarity or distance measures have been proposed and widely applied, such as cosine similarity and the Jaccard correlation coefficient. Meanwhile, similarity is often conceived in terms of dissimilarity or distance as well [Huang, 2008; Taghva, et al., 2010]. Measures such as Euclidean distance and relative entropy have been applied in clustering to calculate the pair-wise distances.

A variety of approaches have been proposed to model document similarity based on different foundations. Some traditional approaches calculate similarity according to document contents (especially document-term relationship), such as Vector Space Model [Salton, et al., 1975], n-gram measures [Damashek, 1995; Kondrak, 2005] and Latent Semantic Analysis [Deerwester, et al., 1990][Dumais, 1994; Landauer, et al., 2007], etc. Recently, by exploiting link structure of objects, some methods focusing on link-based object ranking are proposed by researchers [Xi, et al., 2005; Yin, et al., 2006; Jeh, et al., 2002]. If viewing documents as nodes and relationships among documents as edges, document similarity can be measured by these link-based object ranking methods with the contents of documents ignored.

2.8 PURPOSE OF PRESENT RESEARCH

The purpose of the present research is to propose some methodologies for analyzing the Japanese IT skill standards. These methodologies extract important information that can help human resources during their career to develop themselves. Precisely, the purpose are as follows:

I. Introduces a methodology to analyze the job categories of the Japanese IT skill standards. It introduces a way to sort the required coursed, published by IPA, for each job category.

II. Introduces analysis of the special fields of ITSS. Also, it extracts the required keywords to transfer from the present special field to another one. Also, it clusters the 35 ITSS special fields using K-means algorithm.

III. Introduces a questionnaire that was applied to a Japanese company to grasp the skill of the employees of that company. It presents a comparison between the results of the questionnaire and the results of the analysis of ITSS's documents.

IV. Introduces a methodology to analyze some career paths that were developed by IPA. A document for every level was constructed to can analyze the career paths.

V. Introduces some suggestions for the IT engineers that would like to be a high level project manager or consultant. Moreover, it proposes a typical career path for IT architect.

CHAPTER-3

Japanese Information Technology Skill Standards (ITSS)

3.1 INTRODUCTION

Under the current circumstances where information technology is widely recognized in society as an infrastructure essential for economic activities and people's lives, Japan faces an urgent issue- development of advanced IT human resources that will play a leading role in enhancing the international economic competitiveness of Japan and supporting the healthy development of social systems. At the same time, industry as a whole has become globalized and entered an age of mega-competition that transcends national boundaries. Along with these developments, securing human resources in the IT industry is gradually becoming more difficult.

Advanced IT human resources are defined as people with the sophisticated expertise in IT and other areas who can use that expertise in actual business situations to deliver creative business solutions to resolve issues, generate added value and create business innovation. Such human resources are required to have deep insight and rich experience. As the key people who will support Japan in the future, these resources are required to lead Japan by demonstrating vision, motivating their team members, developing and guiding younger people. furthermore, they must be able to promote the potential exploitation of IT in the embedded software industry mainly in the IT service and manufacture segments as well as general business with employees involved in IT, toward the ultimate goal of contributing to the invigoration of Japan's economy and the improvement of people's lives.

In response to the awareness of this issue, the Ministry of Economy, Trade and Industry (METI) set up in October 2006 the Human Resources Development Group under the information Service and Software Subcommittee of the Information Economy Committee within the industrial Structure Council. In response, METI established ITSS as an industry skill standard in 2002. In July 2007, the working group drafted a report titled Toward

Developing Advanced IT Human Resources . The report summarized Skill Standards for IT Professionals (ITSS), Embedded Technology Skill Standards(ETSS), and User's Information systems Skill Standards (UISS) and defined the correspondence of these standards to the Information Technology Engineers Examination along with a statement that was necessary to build an objective human resources development/ assessment mechanism. Figure 3-1 shows the chronology of the development of ITSS.

METI subsequently transferred the control of ITSS to IPA and entrusted IPA with the administration of these standards. This study concentrate on the analysis of Information Technology Skill Standards (ITSS), to extract important information that help IT engineers to develop their careers, which can help the developers of ITSS to improve the documents of ITSS.

	Period	Output
2001		
2002	Diffusion period	<Dec.> **ITSS V1 Announcement**
2003		<Jul.> ITSS V1.1 Announcement <Jul.> Education Roadmap(6 job categories) (SALES、CONS、ITA、PM、ITS、APS)
2004	Interfusion period	<Jan.> ITSS Introduction Manual Issued <Oct.> ITSS Guidebook Issued
2005		<Dec.> Manager's Guide for ITSS management
2006	Utilization period	<Apr.> **ITSS V2 Announcement** <Oct.> ITSS V2 2006 Announcement
2007		<Apr.> Guide to be an IT Professional <Jun.> Training guideline <Jun.> Guide for in-house Professional Certification
2008～		<Mar.> **ITSS V3 Announcement** <Oct.> **ITSS V3 2008 Announcement**

Figure 3-1: ITSS's Chronology

3.2 ITSS OVERVIEW

3.2.1 Description of IT Skills Standards

The IT Skills Standards are a clear and systematic set of proficiency indicators for the types of skills needed by companies offering various IT-related services. The standards are intended to serve as a common framework to guide the education and training of IT service professionals for business and academia. The IT Skills Standards can be utilized by a wide range of organizations and individuals, including the following:

- IT service providers (including non-IT firms that possess an information systems department): The IT Skills Standards can serve as a guideline to help companies take a strategic approach to the hiring and training of personnel, in accordance with their corporate strategy. Furthermore, companies that have already achieved a certain level of technological proficiency can measure their capabilities against the Standards in order to assess objectively their current level of proficiency.
- Education and training organizations (including institutions of higher Education): The IT Skills standards can provide an objective basis for determining the types of skills that should be taught in education and training programs.
- Individuals (IT professionals): IT professionals could use the IT Skills Standards as a basis for charting out their desired career path, and to determine in what manner they should develop their skills in order to achieve their goals.
- Government: Government agencies could rely on the IT Skills Standards as a basis for formulating support measures aimed at ensuring that IT personnel are trained in an effective manner. With respect to the hiring of new employees, the government could use the Standards as a basis for determining the level of technical proficiency required by its own employees.

The IT Skills Standards are intended to serve as a common framework that is used by various organizations and individuals in order to secure a solid understanding of the whole range of issues related to IT skills.

3.2.2.1 General Versatility of ITSS

Knowledge and skills used in ITSS have certain versatility. Versatility is generality in order to have a certain level of versatility. The knowledge and skill concepts are divided into several layers from the perspectives of both versatility and specificity, as shown in Figure 3-2. Characteristics of knowledge and skills are more general and transferable in lower layers but more definite and specific in upper layers.

ITSS is versatile indices as shown in the middle layers in Figure 3-2. This means skills in ITSS are common levels to various technology areas and business operations.

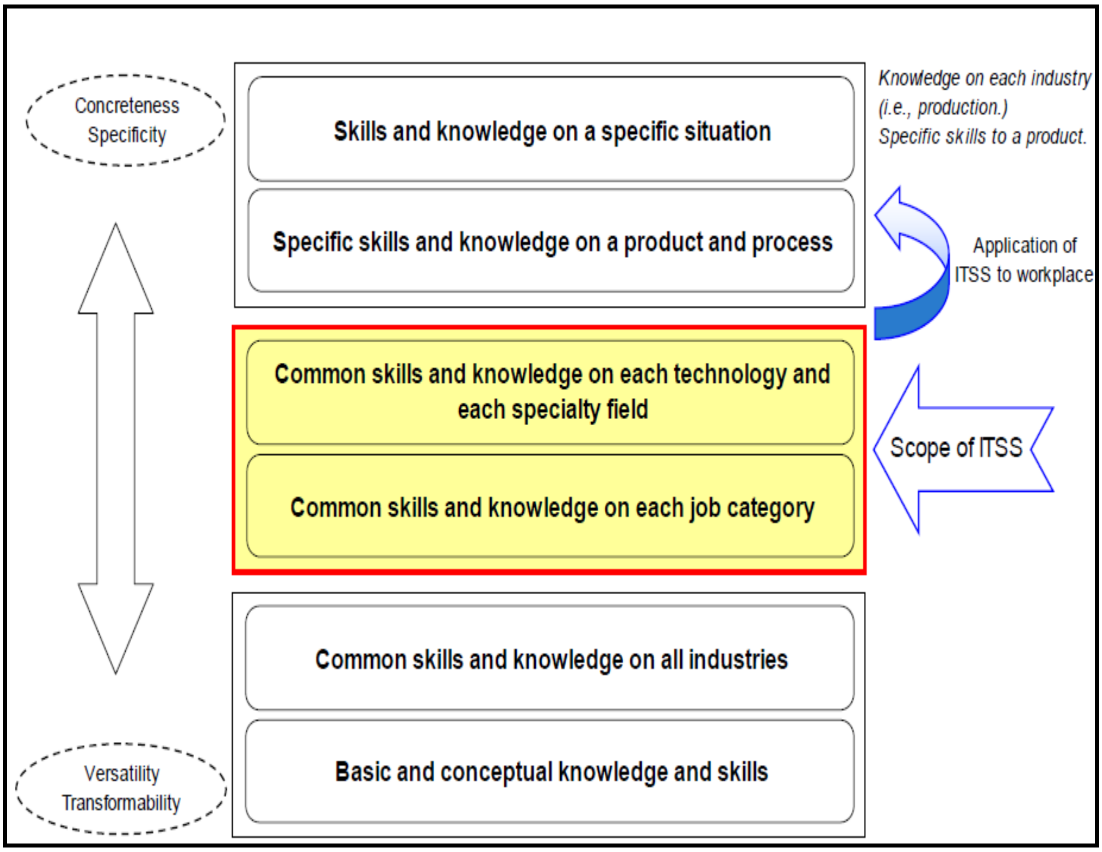

Figure 3-2: Versatility Level of ITSS

3.2.2.2 Issues Not Addressed by the IT Skills Standards

As described above, IT Skills Standards can serve as a common framework for developing the skills of IT personnel. However, simply implementing skill development based on the IT Skills Standards is not sufficient to make companies more competitive, if they fail to make effective use of personnel and skills, or if they fail to adopt a comprehensive business strategy. Clients expect services that meet their requirements in a comprehensive way where the skills of individuals are not always the crux of the issue. Accordingly, companies must engage in internal and external recruiting to secure needed personnel, and must effectively deploy their human resources to provide the services that their clients require. Furthermore, they must also leverage expertise in software engineering to achieve enhanced quality and productivity.

3.3 UTILIZATION OF ITSS

To develop world-class IT human resources, IPA has established the Common Career/Skill Framework to serve as a reference model for ITEE and three skill standards (ITSS, ETSS and UISS). This framework serves as the foundation for developing highly skilled IT human resources. In the previous IT HR market in Japan, there was a typical linear career path (a programmer → a systems engineer → a project leader.) Most IT professionals in Japan are concentrated in the intermediate-level population. With a view to meet diversified and deepening customer requirements with their satisfaction, the IT industry needs to develop expert engineers as IT professionals in both new solutions and latest technology areas. ITSS provides common guidelines for IT markets, which describe direction and contents of a variety of skills.

ITSS's framework is a tool for HR development that organizes the IT services industry into 11 job categories and 35 specialty fields. Every special field has seven levels that depend on the performance and experience of the human resources as shown in Figure 3-3.

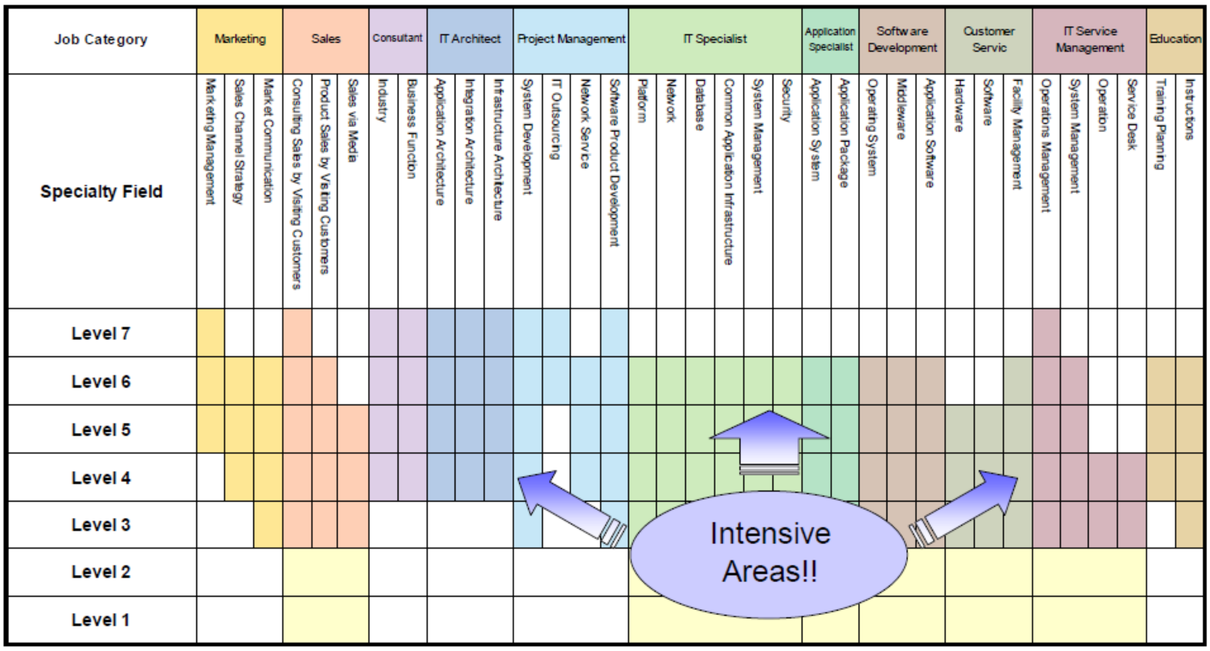

Figure 3-3: ITSS Common Career Framework

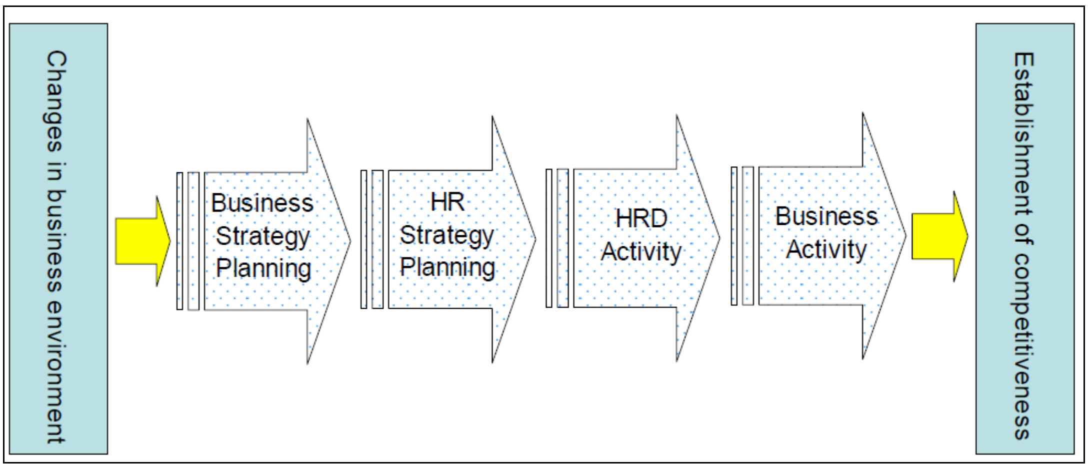

Figure 3-4: Human Resources Investment Process

3.4 ITSS AND HR DEVELOPMENT

HRD without business strategies will not lead to an HRD policy, which develops IT professionals playing vital roles in technology and business for their companies and supporting company competitiveness. ITSS is officious and meaningful when utilized in an HRD process based on a business strategy as shown in Figure 3-4.

Companies ought to define direction and policy by considering their business issues. This is called “business strategy planning” that is the first step to identify and invest company resources. The second step is “HR strategy planning.” which identifies skills necessary to implement business strategies, and determines how to resource companies with skills required for business strategies. In other words, to determine what types of HR are necessitated for business, to analyze skill levels of HR of the moment, and to determine how to fill gaps between them. The third step is “HRD activity,” implementation of an HRD strategy and management of a HRD progress. Finally, the last step is “business activity” to utilize developed HR in actual business to give companies competitiveness advantages.

3.5 TERMINOLOGY AND DEFINITIONS

3.5.1 Skills (Business Capabilities)

Skills in ITSS are defined as business capabilities6. Skills do not mean a set of individual technology elements. Skills are business capabilities to be able to select and apply technology elements appropriately in solving problems.

3.5.2 Professional

An IT professional is a person who actually and successfully achieves business results and contributes to the IT industry growth. Necessary criteria of an IT professional are to be able to,

- Achieve commitments to a customer and company.
- Train and develop subordinates for passing on her/his skills and knowledge acquired through experience.
- Perform activities continuously to improve her/his business capabilities.
- Take social responsibility and have ethics of an IT professional.

An IT professional fulfills customer's requirements to achieve business outcomes by combining and utilizing appropriate skills. Possessing high skills means providing a customer, project members, partners, and an own company with great value that IT professional produces. To achieve a commitment to a customer and a company, IT professional should have not only high technological skills, but also high-level personal skills such as communication, negotiation, leadership, and overall business related skills. Moreover, to pass down skills, IT professional ought to contribute to training and development of subordinates by mentoring, coaching, etc..

3.5.3 Job Category (Career)

Job categories (careers) in ITSS are specialized domains where each professional stands, but they do not precisely indicate roles within a company or project.

3.5.4 Responsible Person

A responsible person is one who has responsibilities in a service provider side and in charge of entire pertinent activity processes and phases. The responsible person delivers value directly to a person in charge in a customer side.

3.5.5 Leader

A leader is one who has responsibilities for promoting execution of particular pertinent activity processes and phases she/he is assigned. The leader delivers value directly to the responsible person.

3.5.6 Member

A member is one who creates deliverables and business outcomes. The member delivers value directly to the leader.

3.6 FUNDAMENTAL STRUCTURE OF ITSS

3.6.1 Structural View of Human Capabilities and Competency

There are two perspectives of human capabilities. One is actual results and the other is capabilities that an individual possesses, as shown in Figure 3-5.

The left side of Figure 3-5 illustrates viewpoints of how many results an individual achieves, how high quality of these results an individual produces is, and how much an individual contributes to business. This figure also explains that an individual's capabilities are assessed by considering a problem solving capability and a performance level at which an individual exercises in assigned duties. The right side of Figure 3-5 illustrates value of an individual, and it shows viewpoints of how much an individual possesses capabilities. From the viewpoint on the left side, there are four essential features of assessment: quality, quantity, collaboration, and responsibility. These four elements are assessed with regard to business performance and outcomes; whether or not quality level is fulfilled, productivity is sufficient, an individual works with others in cooperative and collaborative manner in her/his assigned duties, and an individual achieves business outcomes in a responsible position.

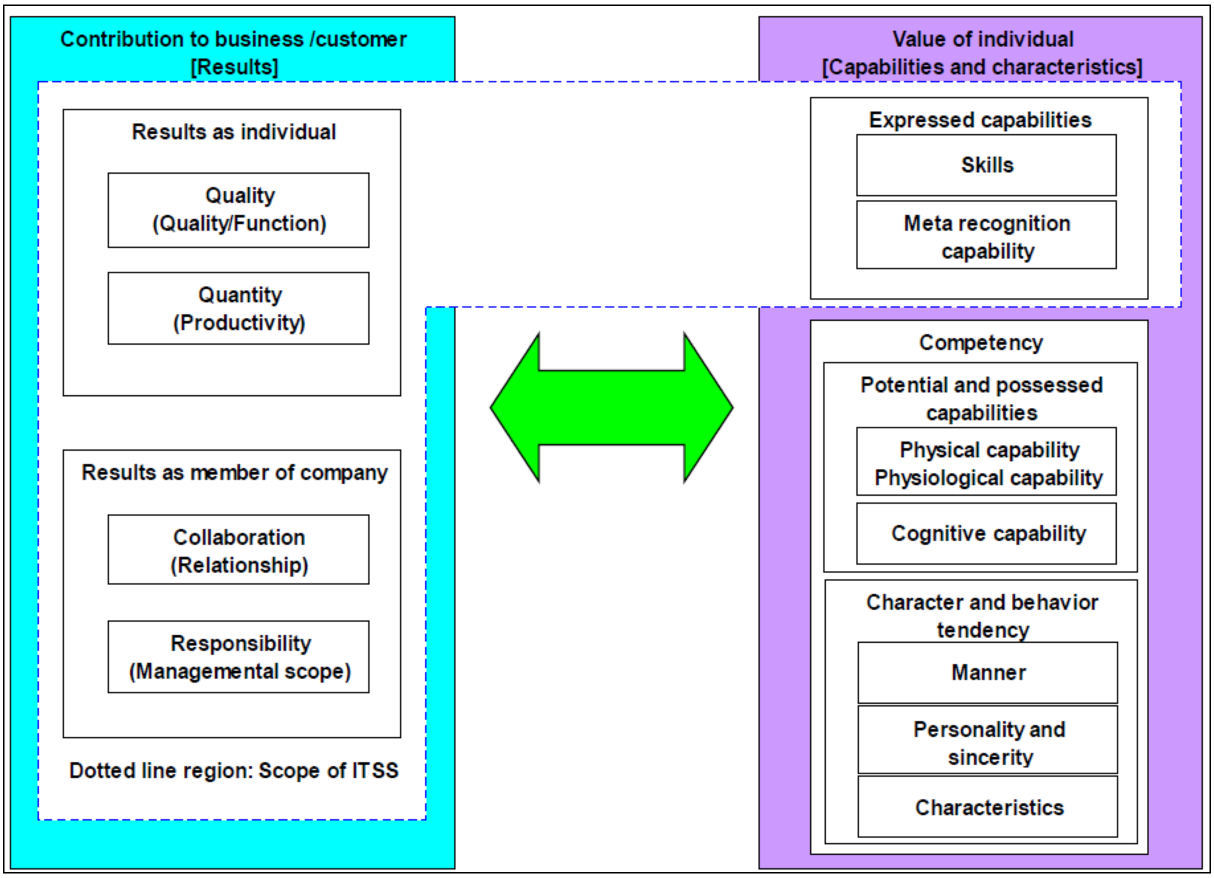

Figure 3-5: HR Investment Process

A difference between viewpoints of left side and right side is explained by a fact that each side is used for assessment of different types of HR elements. The scope on the left side is a “performance assessment,” which depends on a difficulty level of a problem to be solved and magnitude of an outcome. The “performance assessment” allows to judge whether or not an end outcome satisfies a customer and is fruitful for both the customer and an individual’s company. The scope on the right side is a “capability assessment,” which expressed capabilities an individual demonstrates. The “capability assessment” aims to develop desired HR and to ensure steady work performance, which is also useful information at job assignment and personnel allocation in the future.

3.6.2 ITSS Structure and Its Concept

Based on the idea in section 2.6.1, ITSS structuralized its concept as shown in Figure 3-6. In Figure 3-6, problems to be solved are indicated as "Tasks". Contents of tasks vary among jobs, and defined task contents of each job as job categories in ITSS. When an individual performs tasks by utilizing her/his skills, it yields business outcomes.

Utilizing proficient skills, are necessary to take on tasks, in order to put optimal performance that enables her/him to achieve business outcomes that satisfy customer’s demands. A skill demonstration level of an individual has influence on whether or not she/he can achieve a satisfactory result. ITSS outlines assessment of measurable human capabilities in two forms. The first one is skill proficiency levels that indicate capability levels of being able to demonstrate skills. The other one is the KPIs that indicate magnitude of outcomes or results.

ITSS assesses measurable and observable capabilities and skills that are acquirable posteriori. Obtaining necessary knowledge through education or training and using this knowledge in real experience will develop these posteriori skills. This skill learning-demonstration cycle is regarded as experience results. Repetition of this cycle accumulates experience and results, and it directs individuals to a next step.

3.6.3 ITSS Basic Structure

ITSS has two major parts: careers and skills. A career framework is an at-a-glance chart that represents indices of outcomes required for business. A skill dictionary is another at-a-glance chart that lists skills necessary to achieve outcomes. Figure 3-7 shows the structure of ITSS.

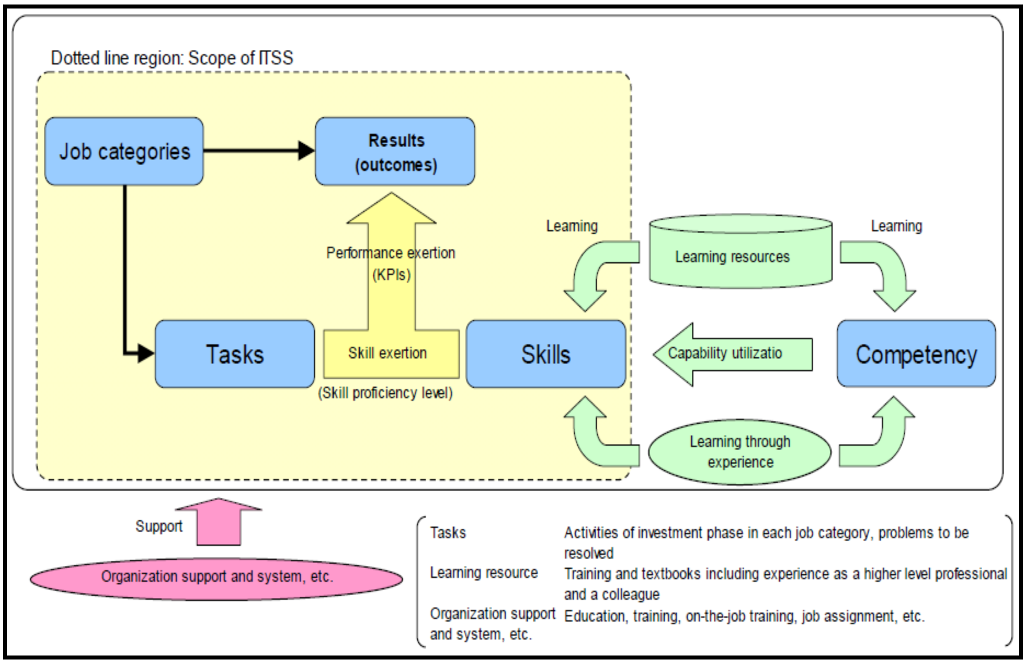

Figure3-6: ITSS Concepts

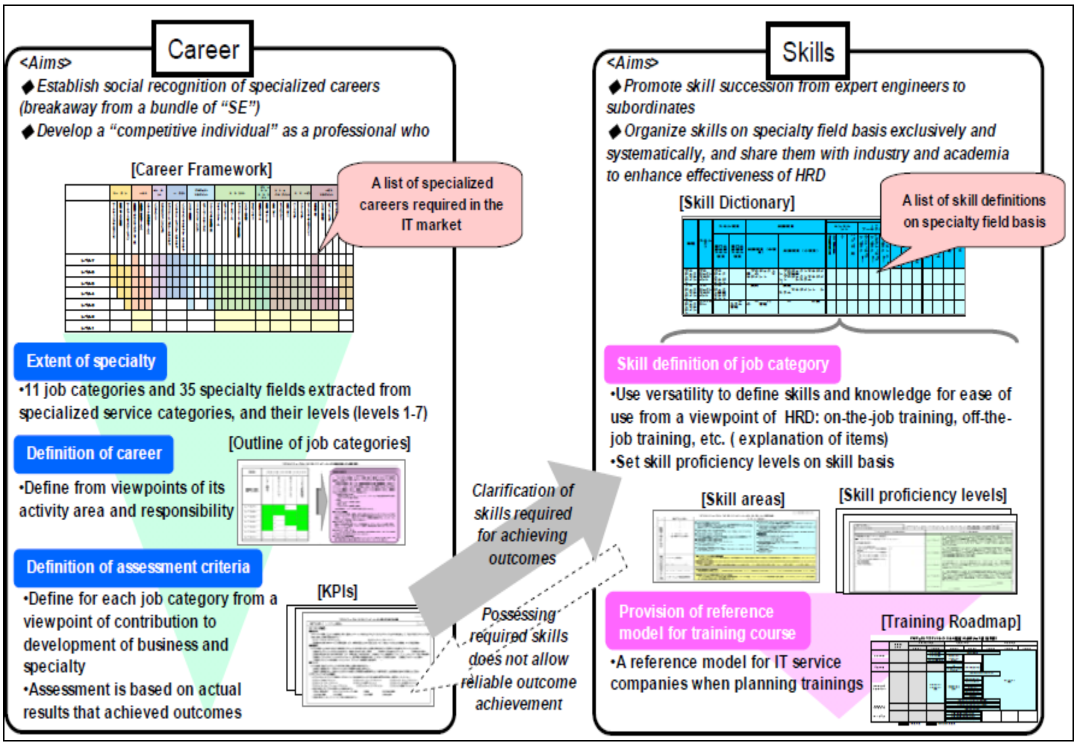

Figure 3-7: ITSS Basic Structure

- Documents on Careers

An activity area of each job category is determined, and this activity area is called a job category. A job category is broken down into a several specialty fields. ITSS sets KPIs of each specialty field in a job category, and the KPIs are objective indices to describe experience and results.

- Documents on Skills

Skills required for each job category or specialty field are broken down to elements called skill items. A skill proficiency level is defined to state a maturity degree of each skill item. Furthermore, each skill item has a set of knowledge items needed for demonstrating the skill item. Skills are organized from viewpoints of both objective observation, and possible use for educational training.

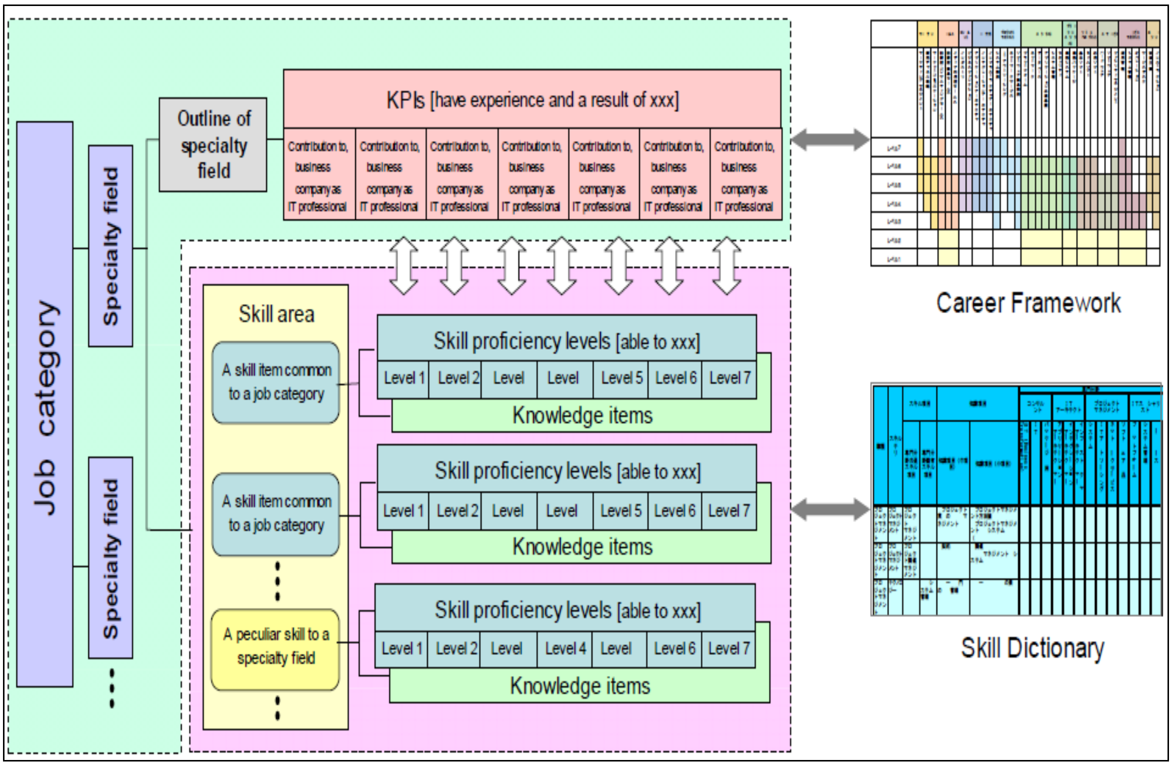

Figure 3-8: ITSS Component

3.6.4 Components of ITSS

Components of ITSS are shown in Figure 3-8. The career framework shows an overall view of ITSS, and it has two axes: the vertical axis exhibits job categories, and the horizontal axis represents the depth of capability levels of the job categories. It provides individuals and companies with a common framework to design, plan skill strategies and career paths. In consideration of business needs, technological specialty and originality, liability to customers, and global recognition. Finally, ITSS defines 35 specialty fields in 11 job categories, and categorizes them into levels 1-7.

- Job Category Outlines

The outlines of the job categories describe activities required of each job category. It is explained that in which occasion and what type of value an individual particular job category should produce at an event of IT investment.

- KPIs (Key Performance Indicators)

The KPIs are defined as metrics of each job category and specialty field, and they are used to identify levels of business capabilities objectively. A distinctive feature of ITSS is assessment of capabilities by the KPIs that are based on actual experience and results. The KPIs consist of two types of contribution. One is business contribution: an individual's direct contribution to an outcome of a project. The other is professional contribution: contribution to a company and the IT service industry as an IT professional. At assessing HR levels, both of these contributions are assessed in a comprehensive manner.

- Skill Dictionary

The skill dictionary is a chart that covers all skill items and knowledge items used in ITSS. The skill items and knowledge items are hierarchically listed in the skill dictionary, and it clearly describes the relation of the skill items and knowledge items to the job categories and specialty fields.

- Skill Items and Knowledge Items

The skill items are definitions of capability elements necessary to achieve business outcomes. The knowledge items are knowledge required for acquiring the skill items. The skill items and knowledge items show their mutual relationship to help ITSS users to identify corresponding knowledge items that an individual should acquire to attain a skill item.

- Skill Areas

The skill areas specify skill items and knowledge items necessary for each job category and specialty field. A description of a skill area consists of common skills to a job category and specific skills to a specialty field in a job category.

- Skill Proficiency Levels

The skill proficiency levels are set on each skill item to indicate skill maturity degrees required for performing business activities. The skill proficiency levels give an explanation of criteria in the term "being able to do" duties, and they are supportive evidence of whether or not an individual possesses skills of a certain level. In other words, an individual with a certain degree of mastery should be able to exercise skill items of a corresponding level of skill proficiency.

- Training Road Map

The Training Road Map illustrates training subjects that individuals ought to complete to the aim at acquiring knowledge of a job category. The Training Road Map is in accordance with ITSS. It consists of "training course groups" and "training course lists" that offer a group of training courses of each job category, "training course details" that describes contents and details of each training course, and "skill items" and "knowledge items" that explain relationship of skill and knowledge items with each training course.

3.6.5 Relationship Among ITSS Components

Figure 3-9 shows the relationship among components of ITSS from perspectives of business outcomes and human capabilities.

- Documents from Business Performance View

As mentioned before, the job categories are shown in the career framework, and there is an overview of each job category in Part2: Career. The overview consists of specialty fields' levels and outlines. In addition, the KPIs are defined for each level of the specialty field.

- Documents from Human Capability View

All skill items defined in ITSS are organized in the skill dictionary. The skill area is a set of skills required for IT professionals to perform business activities, and it is extracted from all skill items. Moreover, the skill area is defined on a job category basis. All skill items necessary to each specialty field are defined with the skill proficiency levels. The skill proficiency levels indicate skill maturity degrees required to achieve levels defined in the KPIs.

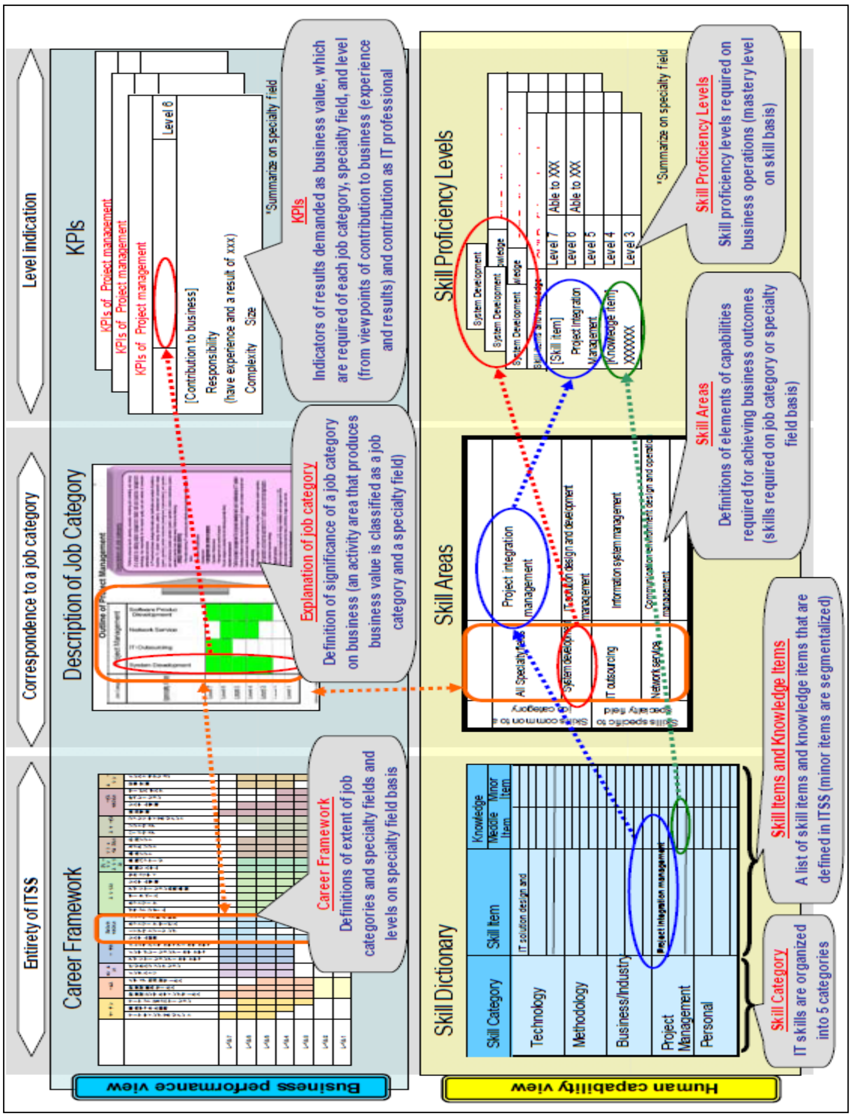

Figure3-9: Relationship Among ITSS Components

3.6.6 Level Assessment by KPIs

The KPIs are assessment indices of individuals as IT professionals while the skill proficiency levels are assessment indices of single skills. Although both have level concepts, remember do not to confuse the KPIs with the skill proficiency levels. The KPIs are based on the proficient skills of the specialty fields which are an assembly of each skill and knowledge. They define levels of experience and results required to create outcomes that external or internal customers demand.

From business success point of view, ITSS takes a particular note on actual value that IT professionals produce through interactions with customers. Elements of practical business capabilities required to produce actual value are broken down and organized systematically in ITSS. Thus, IT professionals are assessed by the KPIs, which are metrics of experience and results. The skill proficiency levels are for assessment of capabilities in a limited range, and they are necessary yet insufficient for assessment of whole business capabilities. This is because improving a proficiency level of individual skill itself is not a final goal of HRD. What ITSS emphasizes is development of IT professionals that can fulfill customers' requests based on their proficient skills.

Levels used in the KPIs have seven stages that represent degrees of problem solving experience and results required of IT professionals in a specialty field in a job category. Figure 3-10 provides a general perspective of 7 levels without considering a particular job category or specialty field.

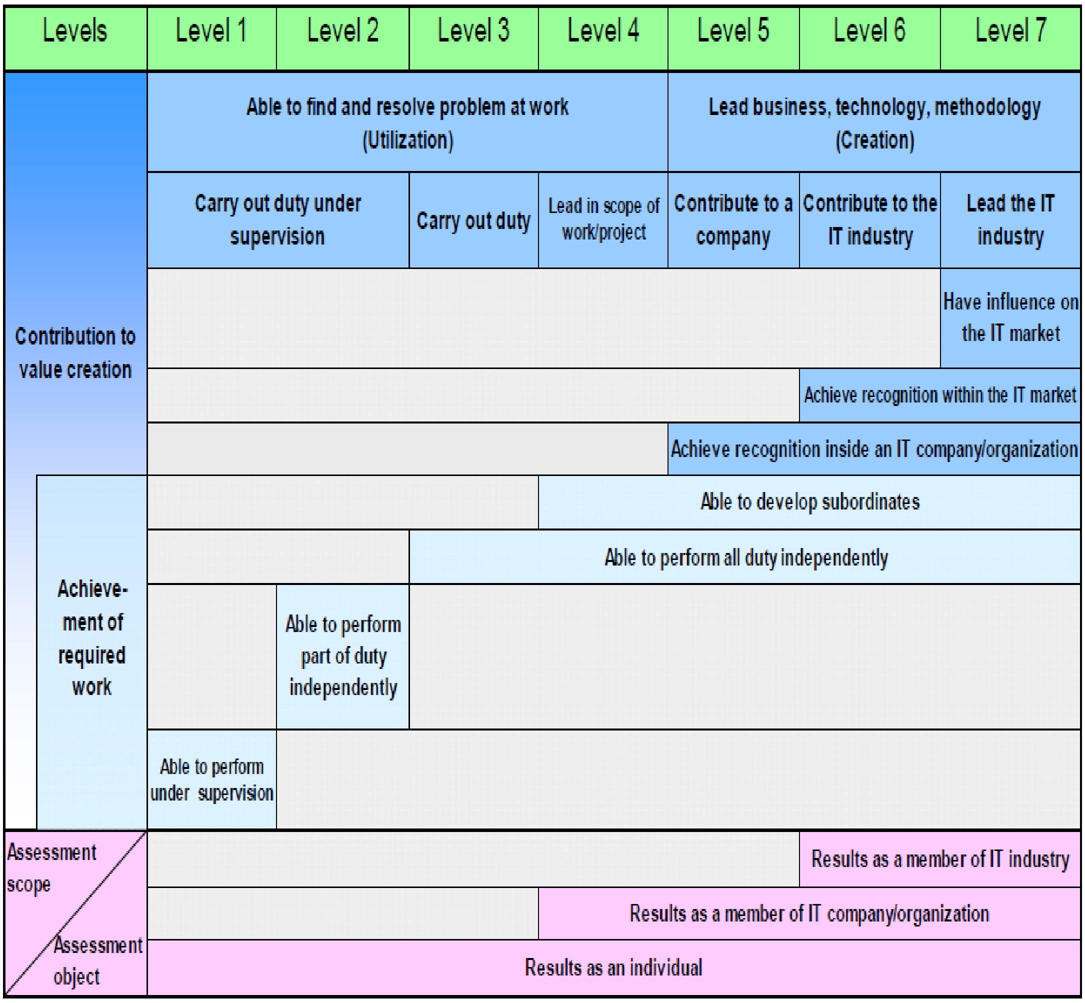

Figure 3-10: ITSS Levels and Assessment

- **Level 7**

Individuals at this level have established their specialty fields as IT professionals, and create and lead technologies, methodologies, and business opportunities inside and outside their companies, organizations, etc.. They have experience and results of leading advanced service development and commercialization for the whole IT market, and are recognized as world-class IT professionals.

- **Level 6**

Individuals at this level have established their specialty fields as IT professionals, and create and lead technologies, methodologies, and business opportunities inside and outside their companies, organizations, etc.. They have experience and results in the IT market besides in their companies, organizations, etc., and are recognized as high-end IT professionals in their countries.

- **Level 5**

Individuals at this level have established their specialty fields as IT professionals, and create and lead technologies, methodologies, and business opportunities inside their companies, organizations, etc.. They have experience and results, and are recognized as high-end IT professionals within their companies, organizations, etc..

- **Level 4**

Individuals at this level have established their specialty fields as IT professionals, and independently clarify problems and lead problem solving with their own skills. In their companies, organizations, etc. they contribute, based on their experience, to organization of knowledge required of IT professionals and to development of lower level IT professionals, and are recognized as high level IT professionals. They are also requested to improve their skills continuously for their career development.

- **Level 3**

Individuals at this level perform all assigned duties independently, and aim to establish their specialty fields. They have application knowledge and skills necessary to become IT professionals, and are requested to improve their skills continuously for career development.

- **Level 2**

Individuals at this level perform assigned duties under the supervision of higher-level

IT professionals. They have basic knowledge and skills necessary to become IT professionals, and are requested to improve their skills actively in order to pursue their career paths for career development.

- **Level 1**

Individuals at this level have the basic knowledge minimum required of those who are relating to IT. They are requested to improve their skills actively in order to pursue their career paths for career development.

There is a wide gap of required capabilities in the KPIs between levels 1-3 and levels 4-7. The main point of the KPIs of levels 1-3 is that "individuals have experience of taking part in projects as team members." This means that an emphasis is on responsibility and performance for their assigned duties; however, it does not contain responsibility or capabilities required of high level IT professionals (e.g., development of lower level IT professionals and successful project conclusion.) Meanwhile, the main point of the KPIs of levels 4-7 is that "individuals have experience of concluding projects successfully as team leaders.".

The individuals at level 3 are requested to "independently perform" duties that "fulfill business requirements." Because they are accountable for outcomes of duties as team members, they are expected to possess "practical business capabilities"; however, they do not need have comprehensive business capabilities such as personal skills including leadership. Levels 1 and 2 of ITSS describe that individuals at these levels are "able to perform" assigned duties "under the supervision of higher level IT professionals." What meant here is to have "knowledge" and "skills" necessary to carry out duties assigned by team leaders with the direction of higher level IT professionals which is requested rather than to have "practical business capabilities" to solve problems. Those employees at level 1 are expected "to do part of duties under the supervision of higher level IT professionals," and a major focus requested at level 1 is on possessing "knowledge."

3.7 ITSS'S ADOPTION PROCESS

Just as companies employ a variety of business strategies, companies invest their capital in different careers and vocations. At an ITSS application to a company, it should first should use definitions of ITSS as common indices and then set its own indices that reflect career and skill requirements specific to a company along with its business strategy. For any organization there are some specific steps have to do, to adopt ITSS in that organization. These steps description are shown in Figure 3-11 as follows:

1. ITSS adoption plan: the organization should determine its objective scope, form its team work, set its activity schedule priority, etc.
2. Job categories definition: this is one of the most complex steps in ITSS's adoption process for any company. The organization should determine the needed job categories and special fields.
3. Level definition: There are seven levels that refer to performance and experience of an IT engineer. The organization should determine the levels of its engineers.
4. KPIs definition: The organization should define the business contribution and professional contribution for every IT engineer in it.
5. Skill proficiency definition: the organization should define the proficiency of the required skills to achieve business outcomes which are selected from the skill dictionary.
6. Assessing KPIs: the KPIs that have been defined in step 4, must be assessed in this step. The business contribution and the professional contributions must be evaluated.
7. Assessing skill proficiency : the company must assess the skill proficiency of every employee using the skill items that were defined by ITSS.
8. Analyzing assessment results: the company must analyze the results of the assessment steps to make a report of personal and management data. Using the assessment results the company can determine the current portfolio of each human resource, to know what are the current status of the company. Then using that current HR portfolio to develop the HR development plans.
9. HRD plans: the company must take into consideration two perspectives, career

development and skill development. It is essential to acknowledge both individuals' current careers and target careers . At the same time, HRD plans should take account that individuals are able to improve skills in a planned and intensive manner.

10. HRD implementation: when implementing HRD, companies to design HRD plans based on actual business practice, which are off-the-job training and on-the-job training.
11. Analysis and evaluation: there are two kinds of assessment; one is experience and results, and the other is technologies specific to a variety of products and methodologies.

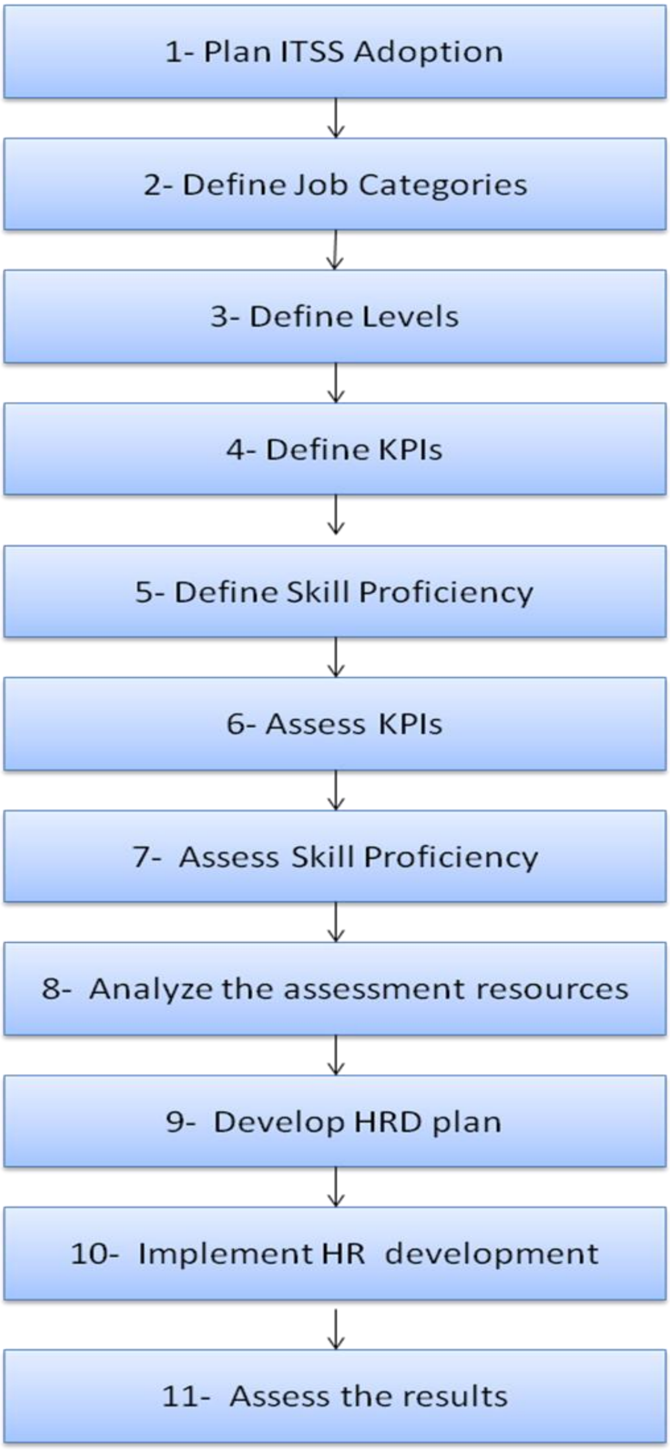

Figure 3-11: ITSS Adoption Process

3.8 ITSS AND OTHER SKILL STANDARDS

This section compares between ITSS and other skill standards in different regions of the world. These skill standards are : e-competence frame work, skill standards for information age (SFIA), and IT skill standards (ITSS).The comparison will be applied according to a specific factors. These factors are: the region, job categories concept, level concept, aim, principle target, users as shown in Table 3-1 .

Aim:

The objective of ITSS is to promote effective capital investment for human resources development in the IT service industry by utilizing the index provided by the skill standards. SFIA's aim is to provide a management tool to help those who are making decisions about the use or development of skills. For this reason the skill definitions are diagnostic, not prescriptive: they contain enough information to enable a rational management judgment as to whether someone has the skill, and in which level. They do not attempt to list all the things that the skilled individual might be able to do. *E-competence's* objective is to create long term HR and competence development strategies for ICT community in EU.

Job Categories:

ITSS's framework has two axes: the vertical axis exhibits job categories, and the horizontal axis represents the depth of capability levels of the job categories. It provides individuals and companies with a common framework to design and plan skill strategies and career paths. While SFIA framework provides a clear model for describing what ICT practitioners and users do. It is constructed as a two-dimensional matrix. The matrix shows a comprehensive set of skills used by both ICT practitioners and users. The two critical dimensions of the SFIA are the skills and levels. In case of E-competence framework, it is based on competence areas and competences, instead of job profiles, being competence-based approaches more flexible and fostering local personalization. The European e-Competence framework is structured from four dimensions. These dimensions reflect different levels of business and human resource planning requirements in addition to job/ work proficiency guidelines.

Skill Concept:

Skills in ITSS are defined as business capabilities. Skills do not mean a set of individual technology elements. Skills are business capabilities to be able to select and apply technology elements appropriately in solving problems. In SFIA, IT professional capability comes from a combination of professional skills, behavioral skills and knowledge. Experience and qualifications validate that overall capability. These skills are defined in a way that makes them easily recognizable in the workplace: the practical nature of the descriptions means that they can effectively be used to describe the capabilities of a person and the needs of a job or class of professionals.

E-competence framework skills are a set of reference e-Competences for each area, with a generic description for each competence. 32 competences identified in total provide the European generic reference definitions of the framework.

Assessment:

ITSS has established a set of key performance indicators (KPIs) that can also help organizations to design their rewards and recognition programs, succession planning, and promotion. There are seven levels from the entrance level to the top class level. The human resource level is judged by his performance and experience. SFIA defines 7 levels of Responsibility, each defined within the context of Autonomy, Influence, Complexity and Business Skills. These 7 levels describe the progression of a practitioner from "Follow", being someone in an entry-level position with no discretion and working under close supervision, through to "Set strategy/inspire/mobilize", being someone with overall responsibility for all aspects of a significant area of work. In E-Competence, for each e-competence suitable proficiency level specifications ranging between e-Competence levels e-1 and e-5 have been constructed. They relate to EQF levels 3 to 8. This dimension involves “behaviors” and levels of autonomy, a bridge from “organizational” to “individual” competences is created. Note that organizational competences are generic and broad whilst individual competences are specific and customized.

Table 3-1: Comparison between ITSS&E-Competence&SFIA

Comparison Factors/Framework Name	ITSS	E-Competence	SFIA
Region	Japan	Europe	United Kingdom
First announcement	2002		2003
Job categories Concept	11 job categories&35 special fields	is structured from four dimensions, Plan-Build-Run-Enable-Manage.	Skills are grouped into subcategories or "business roles". Subcategories are grouped into six categories or work areas.
Level Concept	7 levels according to performance and experience	e-Competence levels e-1 to e-5, which are related to the EQF levels 3 to 8	7 levels- is defined in terms of autonomy, influence, complexity and business skills.
Aim	Development of human resource in high level area.	development of long-term human resources and competence development solutions for the European ICT community.	manage capability such that the right skilled people are in the right place at the right time
Principle Target	IT Engineers with high level business capability in companies	ICT business practitioner in Business Application area	IT Engineers within organizations
User	-IT service companies -Individual IT engineers -IT user companies/organizations	-ICT user and supply companies, ICT practitioners, managers and HR departments, the public sector, educational and social partners across Europe	-Individual IT engineers -IT users and providers companies/organizations

CHAPTER-4

The Analysis of ITSS's Job Categories

4.1 INTRODUCTION

As stated in previous chapters, this dissertation study analyzes Japan's initiative for developing the information technology skill standards. The aim of this chapter is to analyze the activity area of each job category that was determined by ITSS using text mining techniques. We used some of that techniques to measure the similarity function between the different job categories. Moreover, we proposed a method to sort the required training courses list for each job category that were suggested by IPA. As stated in chapter 3 , these training subjects individuals ought to complete to aim at acquiring knowledge of a job category.

In order to improve accuracy, it is necessary to define the weight of the extracted keywords. The higher weighted keywords indicate the significance and priority of learning the associated skills. Keywords with a high weight are important for identifying the required skill for different job categories. So, it is necessary to learn the skills related to the high weight keywords. Using these high weight keywords, the required learning courses for ITSS job categories were sorted. This sorting allows IT engineers to know what are the most important courses have to learn first, so they can concentrate on studying the most needed courses for their career instead of wasting time in studying not important courses for them.

However, the clarification of the required skills for each job category is important for IT human resources, the clarification of the skills required to move between different job categories is also important. The authors have suggested a method that derives the required keywords to move between different job categories.

In this chapter each job category is represented by a text document. As mentioned we have 11 job categories, so we have 11 text documents. These documents are published by IPA. Hereafter, this chapter describes the 11 job categories as shown in Table 4-1.

Table 4-1: Job categories names abbreviations

Categories Names	Marketing	Sales	Consultant	IT Architect	Project Management	IT Specialist	Application Specialist	Software Development	Customer Service	IT Service Management	Education
Abbreviation	Mrk	Sal	Cnslt	IT-Arc	ProMng	IT-Spl	ApSpl	SwDpt	CstSvc	IT-SM	Edu

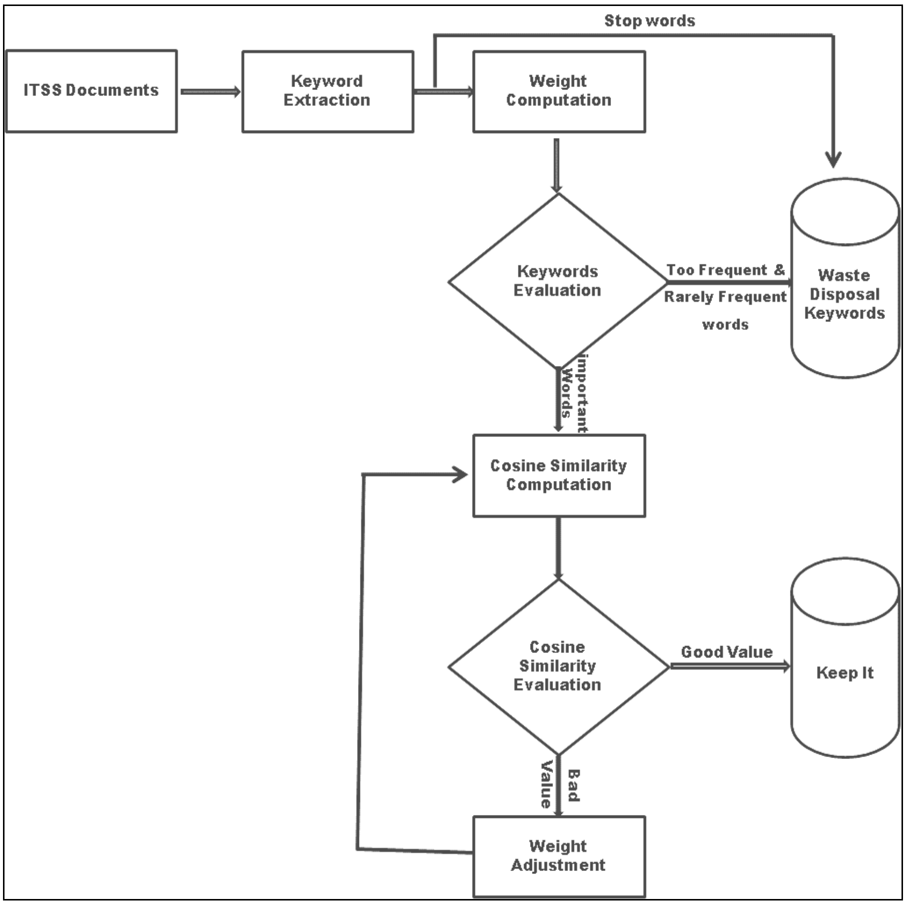

Figure 4-1: Process flow

4.2 PROCESSING FLOW

As stated in chapter 3, ITSS is utilized as a tool for developing professional human resources to implement corporate strategies. Organized into a career framework, ITSS classifies the information services industry into 11 job categories and 35 specialty fields. In each field, there are seven levels based on individual experience and results. The outline of the proposed method is shown in figure 4-1. The input if this process is eleven text documents. In order to analyze these documents, we should carry on the following steps.

4.2.1 Extraction Process

Keywords play a crucial role in extracting the correct information as per user requirements. Everyday thousands of books, papers are published which makes it very difficult to go through all the text material; instead there is a need of a good information extraction or summarization methods which provide the actual contents of a given document, as such effective keywords are a necessity. Since keyword is the smallest unit which expresses meaning of entire document, many applications can take the advantage of it such as automatic indexing, information retrieval, clustering, and classification [Bracewell, et al., 2005]. There are a lot of approaches for keyword extraction [Hulty, 2003; Matsuo, et al., 2004; ZHANG, et al., 2008].

In this research we used AnalogX keyword extraction tool (AnalogX) . AnalogX Keyword Extractor extracts all of the keywords of a webpage, then sorts and indexes them based on of their usage and position; once indexed, you can adjust search engine specific weighting factors and keyword criteria to get the best possible view of how a search engine sees your site. AnalogX Keyword Extractor can load up both local files as well as files off other websites, can work through a proxy, and can have separate configurations for as many search engines as you choose to enter.

4.2.2 Pre-processing step

During processing, the following four rules are applied:

a) If the keyword is a stop keyword then delete it. Stop words are very common words as, prepositions and non-content bearing words. The list of stop words differs from a research to another and they are typically used to filter out nonscientific English words that carry low domain-specific information content. In this research we used the default English stop words list (EnglishStopWords). Table 4-2 shows part of the removed stop words.
b) Truncate suffixes and trailing numerals so that words having the same root are collapsed to the same word for frequency counting.
c) If the frequency of the keyword=0 in only two categories or less then delete this keyword. This rule filtered the set of keywords from the words that appeared so frequent in most documents.
d) If the frequency of the keyword=0 in nine categories or more then delete this keyword. This rule removes the keywords that are rarely appearing in the documents.

In Table 4-3 some of the keywords deleted after applying rule c and rule d in the pre-processing step. The ”No of Zero” column expresses numbers of the documents do not contain a specific keyword. As an example, the word ”data” appears in ten documents. This means that this word is so frequent. So, we deleted it after applying rule c. Also, the word ”accident” appears in only one category. So, the word ”accident” is deleted after applying rule d because it appears in only one document (IT-Spl document).

Table 4-2: Part of removed stop words

Keyword	*Mrk Fre*	*Sal Fre*	*Cnsl Fre*	*IT-Arc Fre*	*Pro Mng Fre*	*IT-Spl Fre*	*Ap Spl Fre*	*Sw Dpt Fre*	*Cst Svc Fre*	*IT-SM Fre*	*Edu Fre*
About	0	4	0	0	2	38	6	3	5	0	0
All	0	16	43	0	0	0	0	0	3	0	2
Among	0	0	0	0	0	2	0	0	0	0	0
And	0	294	303	523	652	594	233	322	258	553	144
Annual	0	0	80	0	65	0	0	0	0	0	0
Are	0	2	0	0	2	2	2	2	2	2	0
Area	42	39	32	58	76	130	50	79	99	72	23
Around	0	0	0	12	0	0	0	0	9	0	0

Table 4-3: Part of the keywords deleted after applying the above two rules

keywords	*Mrk Fre*	*Sal Fre*	*Cnsl Fre*	*IT-Arc Fre*	*Pro Mng Fre*	*IT-Spl Fre*	*Ap Spl Fre*	*Sw Dpt Fre*	*Cst Svc Fre*	*IT-SM Fre*	*Edu Fre*	*No of zero*
Abroad	0	0	0	0	0	0	0	0	0	0	2	10
Accessibility	0	0	0	0	0	0	0	3	0	0	0	10
Accident	0	0	0	0	0	2	0	0	0	0	0	10
accounting	0	0	6	0	0	0	1	0	0	0	0	9
Data	1	0	8	3	1	10	4	3	7	7	1	1

4.2.3 Mathematical Representation

There are several ways to model a text document. For example, it can be represented as a bag of words. These words are extracted from the texts of the documents to be used for content identification. This paper used the vector space model to represent the text documents in n-dimensional space. Each individual document D_w is represented as a vector of terms as shown in equation 4-1.

$$D_w = < w_1, w_2, ..., w_m > \qquad (4\text{-}1)$$

And each term w_i in a document D_w is assigned a weight w_i which represents its importance. . The weight of a word expresses the importance of each word in every document. Several terms weighting schemes have been proposed to compute the importance of a term in a document [Buckley, et al., 1995; Jones, et al., 1997; Salton, et al., 1988]. These weighting schemes assign a value to keywords based on how useful they are likely to be in determining the relevance of a document. . In this research the weight of the keywords was computed using equation 4-2,

$$w_i = \frac{tf_i}{\sum_{i=1}^{m} tf_i} \qquad (4\text{-}2)$$

where, tf_i is the frequency of the i^{th} term in document D, m is the number of keywords in document D. This weighting formula measure the importance of a keyword with respect to all the other or the remaining keywords by dividing how many times that keyword appears in a documents by the summation of how many times all the other keywords appears in the document.

4.2.4 Similarity Measure

As we mentioned former in chapter 2, the similarity measure is a function which computes the degree of similarity between a pair of text objects. In this research we used cosine similarity measure to compute the closeness between every pair of documents. The cosine similarity between D_i and D_j is defined by the angle between their feature vectors which are in our case w_i as shown in Equation 4-3.

$$sim(D_i, D_j) = \frac{D_i \times D_j}{\|D_i\| * \|D_j\|} \qquad (4\text{-}3)$$

Where, "×" denotes the dot product of the two vectors D_i and D_j, $\|D_i\|$ denotes the length or norm of a vector D_i.

4.3 TRANSFERRING BETWEEN JOB CATEGORIES

Companies and organizations operate in an increasingly competitive environment, which drives a need for continuous employee skill development. To be successful, workers must make themselves as valuable as possible to be hired by companies. So, the clarification of the required skills and abilities for every job category is important for IT human resources. In addition, the clarification of the required skills to move between different job categories is very important. In our study we proposed a method for deriving the required keywords to transfer from a job category to another. The main issue in this method is the keywords of each job category. Any two job categories have two types of the keywords: Common keywords and Special keywords. Where Common keywords are the words that appeared in the both job categories, while, the special keywords are the words concerned to only one of the two job categories. Every type of the keywords has its weight formula as shown in equation 4-4 and 4-5. The weight $S.W_k$ of special keywords K is computed using equation 4-4.

$$S.W_k = \log(tf_{i,k}, average(tf_{i,1}, tf_{i,2}, ..., tf_{i,n})) \qquad (4\text{-}4)$$

Where, the special keyword K is in the document i, $k = 1,2,...,n$ and the keywords are in the document $i = 1,2,...,11$ (one of the 11 text documents used).

The weight of the common keywords is computed by equation 4-5:

$$C.W_{i,k} = wt_{i,k} * P(wt_{i,k}),$$

$$C.W_{j,k} = wt_{j,k} * P(wt_{j,k}) \qquad (4\text{-}5)$$

Where, $wt_{i,k}, wt_{j,k}$ the weight of the word K in the documents i, j respectively, $P(wt_{i,k})$, $P(wt_{j,k})$ are the probability of $wt_{i,k}$ and $wt_{j,k}$ respectively as shown in eequations 4-6, 4-7 and 4-8.

$$P(wt_{i,k}) = \frac{wt_{i,k}}{wt_{i,k} + wt_{j,k}} \qquad (4\text{-}6)$$

$$P(wt_{j,k}) = \frac{wt_{j,k}}{wt_{i,k} + wt_{j,k}} \qquad (4\text{-}7)$$

$$P(wt_{i,k}) + P(wt_{j,k}) = 1 \qquad (4\text{-}8)$$

Now, for any two documents i and j, they have two sets of keywords: special keywords set and common keywords set. The required keywords to move from document i to document j are:

1. The special keywords for document j and,
2. The common keywords, for document j, those have . $R \geq \alpha$ Where $R = \frac{C.W_{j,k}}{C.W_{i,k}}$, α is a threshold.

For more explanation, figure 4-2 shows the process of transition from document D_1 to D_2.

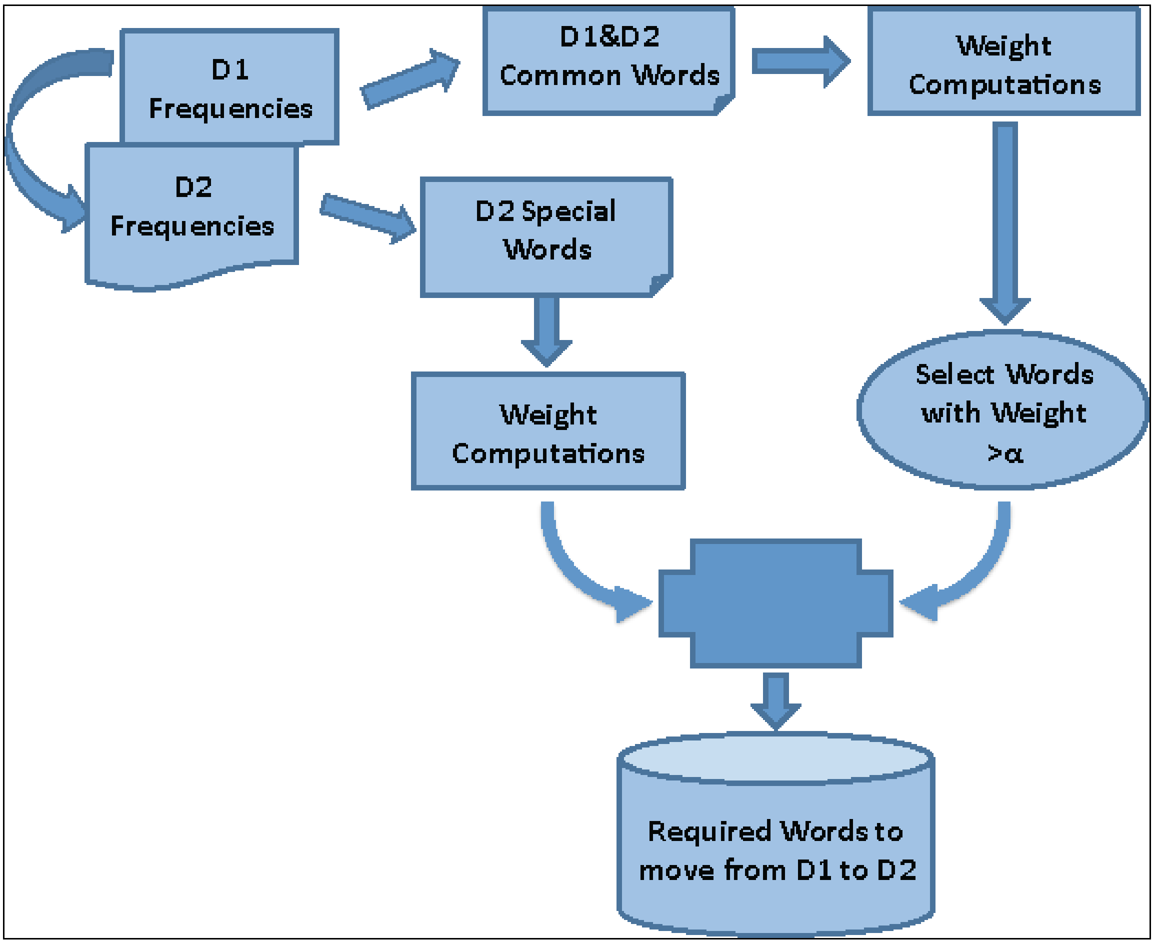

Figure 4-2: D1 to D2 Moving Process

4.4 EXPERIMENT AND RESULTS

For this experiment the proposed method was applied on eleven text documents. Each document represents one job category of ITSS. These documents were published by IPA. The first step is the keywords extraction process. To extract key words, AnalogX keyword extraction tool has been used to extract the keywords from every document. The number of extracted keywords is 449 keywords. This set of extracted keywords contains many repeated words and a lot of stop words and so on. So, it is necessary to perform the pre-processing step. In this step we applied 4 rules.

- Firstly, is the deleting the stop words manually using the default English stop words list.
- Secondly, is the reducing of all the keywords to their roots.
- Thirdly and fourthly, is deleting of the frequent and rarely appeared words.

After applying this step, we had 83 keywords. The third step is the weight computation. We used Equation 4-2 to compute the weight of keywords. Now, the documents are ready to be mathematically represented using vector space model. At this point every document is represented by 83 dimensions vector. Finally, the cosine similarity was computed. Data from Table 4-4 show the result of cosine similarity.

Table 4-4: Cosine Similarity Results

Job Categories	*Mrk*	*Sal*	*Cnsl*	*IT-Arc*	*Pro Mng*	*IT-Spl*	*Ap Spl*	*Sw Dpt*	*Cst Svc*	*IT-SM*	*Edu*
Mrk	1	0.714	0.519	0.018	0.068	0.086	0.036	0.059	0.249	0.191	0.073
Sal	0.714	1	0.768	0.016	0.03	0.029	0.009	0.014	0.039	0.027	0.014
Cnsl	0.519	0.768	1	0.098	0.142	0.106	0.139	0.032	0.05	0.074	0.081
IT-Arc	0.018	0.016	0.098	1	0.225	0.265	0.449	0.306	0.091	0.092	0.022
ProMng	0.068	0.03	0.142	0.225	1	0.459	0.493	0.355	0.358	0.227	0.024
IT-Spl	0.086	0.029	0.106	0.265	0.459	1	0.803	0.418	0.336	0.543	0.146
ApSpl	0.036	0.009	0.139	0.449	0.493	0.803	1	0.562	0.391	0.585	0.092
SwDpt	0.059	0.014	0.032	0.306	0.355	0.418	0.562	1	0.546	0.272	0.074
CstSvc	0.249	0.039	0.05	0.091	0.358	0.336	0.391	0.546	1	0.564	0.011
IT-SM	0.191	0.027	0.074	0.092	0.227	0.543	0.585	0.272	0.564	1	0.104
Edu	0.073	0.014	0.081	0.022	0.024	0.146	0.092	0.074	0.011	0.104	1

4.4.1 Example for the Transition between Job Categories

To explain the results for deriving the required keywords to move between different job categories, the two job categories ProMng and IT-Spl are used as an example. Equation 4-2 was applied to compute the weight of the keywords and then used that weight to represent ProMng and IT-Spl mathematically. ProMng and IT-Spl are two vectors with 83 dimensions. By using equation 4-3 the cosine similarity between the two documents is 0.459. The required keywords to move from ProMng <–> IT-Spl are shown in Table 4-5. Equations 4-4 and 4-5 were used for computing the weight of special and common words weight. Table 4-5 shows special words, common words and the required keywords to move between ProMng and IT-Spl.

For more declaration Figure 4-3 shows the required keywords to move between some special fields as Mrk, Sal, Cnsl, ProMng, appSpl, IT-Spl. Figure 4-3 is composing of 3 blocks: B1, B2,B3 according to the similarity values. The blocks are B1=(Mrk, Sal, Cnsl) , B2=(ProMng, appSpl,IT-Spl) and B3=(Edu). The distance between the job categories in the same block is smaller than the distance between job categories in different blocks. This because the cosine similarity values between the job categories inside any block is bigger than the values of the cosine similarity between the job categories in different blocks Also, according to the similarity results the job category 'Education' is far from all the other job categories.

Table 4-5: Transformation between ProMng and IT-Spl

ProMng S.W	ProMng wt *1000	IT-Spl S.W	IT-Spl wt *1000	ProMng & IT-Spl C.W	ProMng wt *1000	IT-Spl wt *1000	ProMng-->IT-Spl	IT-Spl->ProMng
Adaptation	0.71	Measure	1.44	assessment	0.12	3.02	Measure	Adaptation
Corporate	1.42	Transaction	1.44	consultant	13	14.9	Transaction	corporate
Equipment	1.42	Migration	2.17	definition	22.6	1.46	Migration	Equipment
Intellectual	1.42	advice	4.33	Design	2.2	18.2	assessment	intellectual
Symposia	2.13	operating	4.33	dictionary	0.35	0.36	advice	symposia
Solution	4.96	Internet	5.05	engineering	0.69	18.1	operating	solution
Structure	7.09	regulation	5.05	improvement	3.14	1.1	Internet	Structure
Execution	12.8	science	5.05	Inspection	0.70	0.73	regulation	execution
Administrate	14.2	Structure	5.05	installation	53.3	4.05	science	Administrate
Performance	14.2	architecture	7.94	ITEE	0.35	0.36	Structure	Performance
Verification	28.3	accurately	8.66	maintenance	3.82	14.8	architecture	service
Estimating	42.5	preparation	8.66	manage	6.32	6.55	accurately	definition
Qualitie	46.0	Computer	11.54	method	1.56	7.42	preparation	verification
Activitie	56.7	collaboration	12.99	network	26.0	32.8	Computer	Estimating
Control	96.4	Test	14.43	peak	62.2	62.9	support	qualitie
Contract	119	guidance	20.20	policy	1.41	1.46	collaboration	installation
		scale	23.81	Practice	8.26	9.28	Test	Activitie
		Technologie	25.97	product	4.75	18.7	engineering	control
		Infrastructure	26.70	security	0.04	44.8	Design	Contract
		Platform	33.19	service	15.7	0.11	guidance	
		Specialist	57.72	Software	40.1	12.4	scale	
		responsibility	101.7	successfully	48.1	1.53	Technologie	
				support	0.14	11.8	Infrastructure	
				System	4.21	94.1	Platform	
							security	
							Specialist	
							System	
							responsibility	

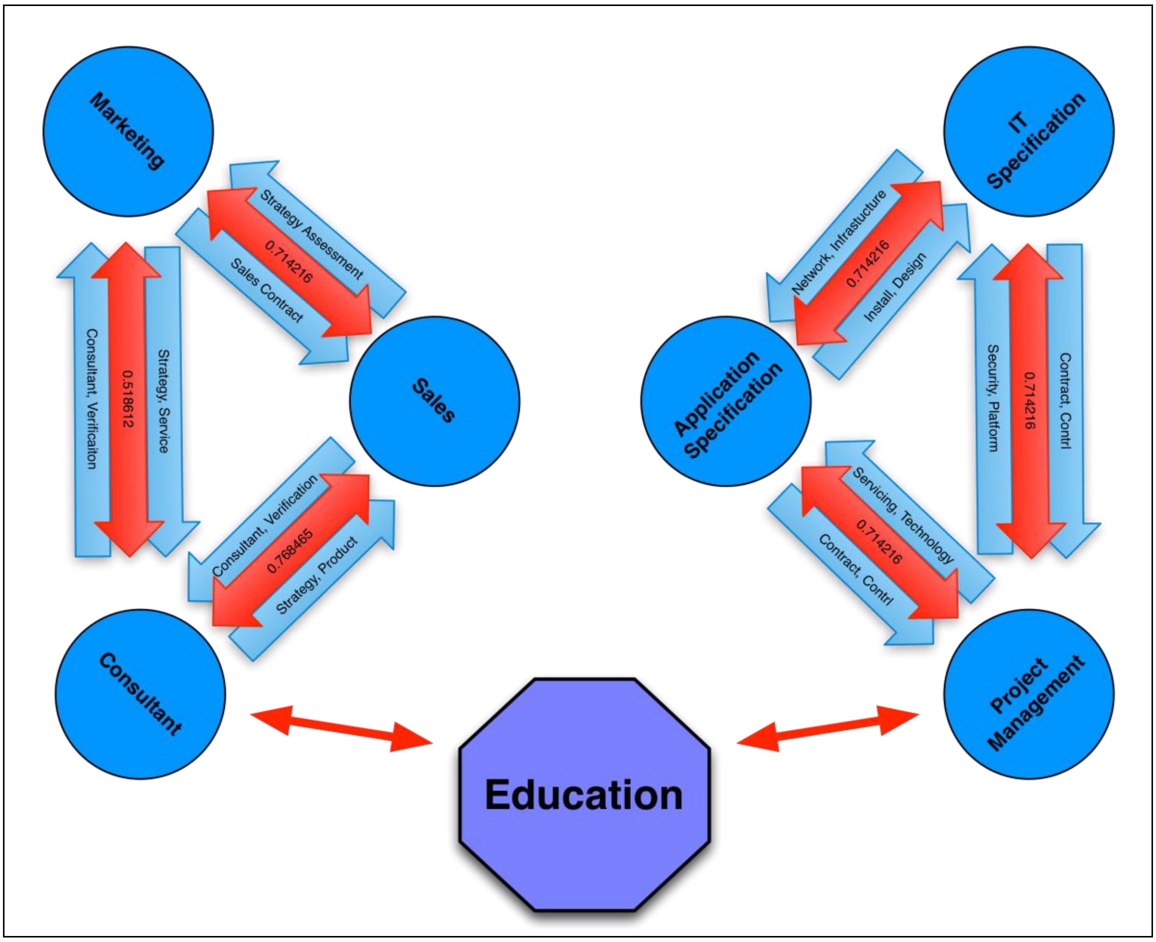

Figure 4-3: Transferring between different Job Categories

4.4.2 High weight keywords and the required learning

The key point of the author's analysis here is the keywords. The keywords were used to represent the text documents to can apply text mining methods. Now, the keywords will be used to sort the required courses for each ITSS job category. But, here only high weight keywords will be used. The higher weighted keywords indicate the significant and priority of learning the associated skills. So, in this research the high weight keywords were used to estimate the most important learning courses for each job category by a very simple method. The required courses for job categories, published by IPA, were used in this method.

Table 4-6 contains the higher weighted keywords in each job category. These keywords have the highest weight in each document. The authors noticed that these keywords are important in realty for their job categories. For example, the words" products, sales, and strategy" are very important for Sal and Mrk jobs.

Table 4-6: The keyword list of the high weight for every job category

Job Category	High Weight Keywords						
Mrk	Sale	Strategy	Product	Segment	Methodology	Service	company
Sal	Assign	Company	Product	Sales	Segment	Strategy	Solution
Cnsl	Consultant	User	Maintaince	Quality	Sales	Definition	Strategy
IT-Arc	Architect	Definition	Design	Maintaince	Peak	User	Solution
ProMng	Peak	Contract	Control	Definition	Manage	Quality	Maintaince
IT-Spl	Maintaince	Platform	Network	Responsible	Security	Specialist	System
ApSpl	Design	Maintaince	Peak	Technology	Security	Specialist	System
SwDpt	Software	Design	System	Peak	Product	Operating	Middleware
CstSvc	Hardware	Install	Maintaince	Segment	Service	Software	System
IT-SM	Maintaince	Network	Security	Service	Strategy	Support	User
Edu	Practice	Methodology	Assign	Company	Goal	Implement	Infrastructure

4.4.3 Proposed method to sort learning courses

The following steps outline the proposed method used to sort the learning courses using highest weighted keywords. The weights are categorized using inputs and outputs using a sex steps process:

- Input: High weight words (K), ITSS courses names(C), Common skills to job category(S).
- Output: Sorted Courses.
- The process includes:

1. For each job category, search if K exists in C , S .
2. If exist, count how many times appears.
3. If not, ignore.
4. Now, every K has a value (how many times appears in C , S).
5. Sum the keywords values for each C .
6. Finally, Sort the courses according their values in step 5.

At the completion of step 5 every course had a value. This value (point) indicates how importance the course is. So, the course with high point should be learned first because this course is important for this job. The obtained result in this research is very important for human resources to determine education recommended for their careers.

Table 4-7 shows some of the required courses for Application Specialist after sorting them. Here there are seven high weight keywords. Every keyword has had a value corresponding to each course. Also, each course name had a value by summing the values of all keywords belonged to this job category. According to this value, the courses were sorted. The course with high point seems to be an important course. Therefore, it is good to learn from courses have high point. For ApSpl job category, the courses, Security system component Technology, Application system design, Distributed computing system component Technology, System Design Fundamentals, should be learned first.

Table 4-7: Required courses for appSpl after sorting

Course Name/Highest weight keywords	Design	Maintaince	Peak	Technology	Security	Specialist	system	Total
Security system component Technology	0	0	0	29	9	0	4	42
Application system design	25	0	0	6	0	0	7	38
Distributed computing system component Technology	1	0	0	37	5	0	3	35
System Design Fundamentals	22	0	0	7	0	0	5	34
Application system operations /Maintaince	0	1	0	11	3	0	11	25
System Operations/Maintaince	0	1	0	11	4	0	7	23
System Management infrastructure component technology	0	0	0	17	0	0	8	23
Database Component Technology	0	0	0	18	1	0	4	23

CHAPTER-5

The Analysis of ITSS's Special Fields

5.1 INTRODUCTION

In this chapter, text mining techniques were used to analyze the documents of ITSS. To be more specific, a clustering algorithm will be applied to classify documents defined by the ITSS. In order to improve accuracy, it is necessary to define the weight of the extracted keywords. We used the TF-IDF weighting scheme to compute the weight of the extracted keywords [Salton, et al., 1988]. The term-frequency (tf) is a document specific local measure and is not the actual term frequency but it is, the actual term frequency divided by the summation of the frequencies of all keywords in that document. This frequency measure is used as the normalization factor because longer documents tend to have more terms and higher term frequencies and its measure, measures how much that term is represented inside the document. The term idf of the weighting scheme is an Inverse Document Frequency measure. The main heuristic behind the idf measure, is that a term that occurs infrequently is good for discriminating between documents.

However, the clarification and systemization of skills and abilities required for each field in IT is very important for human resources, the clarification of skills required for the transition between different fields is also very important. This approach proposed a method that extracts the required keywords to move from field to another. In addition, important keywords, which has a high weight, were determined to may be used to estimate the required skills for IT engineers.

5.2 RESEARCH OF ITSS AND DISTANCE COMPUTATION

This section focuses on clustering of ITSS special fields. Then making a skill-up road map that determined the study or education (skills/abilities) required for changing the career path of any human resource or which skills required to move from job to another job. In this research, each special field is represented by a text document. As mentioned this study has 35 text documents for the 35 special fields. These documents are published by IPA .

Hereafter, this paper describes the 35 special fields as “Marketing Management” with ‘M1’, “Sales Channel Strategy” with ‘M2’, “Market Communication” with ‘M3’, “Consulting

Sales by Visiting Customers” with ‘S1’, “Product Sales by Visiting Customers” with ‘S2’, “Sales via Media” with ‘S3’, “industry” with ‘C1’, “Business Function” with ‘C2’, “Application Architecture” with ‘ITA1’, “Integration Architecture” with ‘ITA2’, "Infrastructure Architecture" “Education” with ‘ITA3’, "System development" with ‘PM1’, "IToutsourcing"with‘PM2’ , "Network services ” with ‘PM3’, "Software product development " with ‘PM4’, "Platform" with ‘ITS1’, "Network " with ‘ITS2’, "Database ′with ‘ITS3’, "Common application infrastructure ′with ‘ITS4’, "System Management " with ‘ITS5’ "Security ” with ‘ITS6’, "Application systems ” with‘AS1’, "Application packages " with ‘AS2’, "Operating system " with ‘SD1’, "Middleware " with ‘SD2’, "Application software” with ‘SD3’, "Hardware " with ‘CS1’, "Software ” with ‘CS2’, "Facility management " ‘CS3’, "Operations Management " with ‘ITSM1’, "System Management ” with ‘ITSM2’, "Operation ” with ‘ITSM3’, "Service desk ” with ‘ITSM4’, "Training Planning " with ‘E1’, "Instructions ” with ‘E2’ .

5.2.1 Processing Flow

The outline of the process is shown in Figure 5-1. The input of the process is 35 text documents. In order to apply clustering algorithm on text documents, some steps should be done as follows:

a) Keyword extraction

Keyword extraction is an important technique for document retrieval, document clustering, summarization, text mining, and so on. By extracting appropriate keywords, the dimension of documents will decrease and can be mathematically represented. There are many algorithms to extract keywords from a document. In this thesis, we use AnalogX keyword extraction tool.

b) Pre-processing step

Initially, very common words, (prepositions and non-content bearing words, often known as stop words) are removed completely from the extracted keywords. Stop words lists are typically used to filter out non-scientific English words that carry low domain-specific information content. This research used the default English stop words` list. Usually these words are eliminated from extracted keywords set because they return vast amount of unnecessary information. In addition, stemming is applied. Word stemming is used to truncate suffixes and trailing numerals so that words having the same root (e.g., activate, activates, activation, and active) are collapsed to the same word for frequency counting. This work applied Porter stemming tool [Porter, 1997]. Now, the documents are ready to be mathematically represented.

c) Mathematical Representation

In our study, the vector space model is used, where each document is represented by The weight (w_{ij}) vector, D_w, in the multidimensional space of document words. The weight of a word expresses the importance of each word in every document. These term's weighting schemes assign value to keywords based on how useful they are likely to be in determining the relevance of a document. This section used the TF-IDF weighting scheme to compute the weight of the extracted keywords. The term-frequency (tf) is a document specific local measure and it is not the actual term frequency but it is, the actual term frequency divided by the summation of the frequencies of all keywords in that document. This frequency measure is used as the normalization factor because longer documents tend to have more terms and higher term frequencies and also this frequency measure, measures how much that term is represented inside the document. The *idf* part of the weighting scheme is an Inverse Document Frequency measure. The main heuristic behind the idf measure, is that a term that occurs infrequently is good for discriminating between documents. Thus the vector space model of a text data set is considered to be a word-by-document matrix whose rows are words and columns are document vectors.

$$D_w = < w_1, w_2, ..., w_m >$$

Where w_i is the weight of the i^{th} term in the document D, m is the number of keywords in D,

$$w_i = \frac{tf_i}{\sum_{i=1}^{m} tf_i} * \log(\frac{N}{n}) \qquad (5\text{-}1)$$

Where tf_i is the frequency of the i^{th} term in the document D, N is the total number of documents in the collection, and n is the number of documents contain the term.

Formula (5-1) used $tf_i \Big/ \sum_{i=1}^{m} tf_i$ instead of using tf. The term $tf_i \Big/ \sum_{i=1}^{m} tf_i$ measures how much the i^{th} term in the document D is represented in that document. But tf counts how many times the term appeared in the document only. So, by multiplying, the value which represent a term in a document ($tf_i \Big/ \sum_{i=1}^{m} tf_i$), by, the value that represent the same term among all the documents **log** (*N/n*), then the obtaining value is good to express how much this term is important for the document.

d) Similarity measure

Document vectors are now ready for clustering, where similar documents are grouped together in the same cluster. A variety of similarity or distance measures have been proposed and widely applied, such as cosine similarity and the Jaccard correlation coefficient. Meanwhile, similarity is often conceived in terms of dissimilarity or distance as well . Measures such as Euclidean distance and relative entropy have been applied in clustering to calculate the pair-wise distances. This research used cosine similarity to compute the closeness between every pair of documents. The cosine similarity of two documents D_i , D_j is defined by the angle between their feature vectors which are in our case w_i as follow:

$$sim(D_i, D_j) = \frac{D_i \times D_j}{\|D_i\| * \|D_j\|} \qquad (5\text{-}2)$$

Where "$\times$" denotes the dot-product of the two vectors D_i and D_j and $\|D_i\|$ denotes the length or norm of a vector D_i .

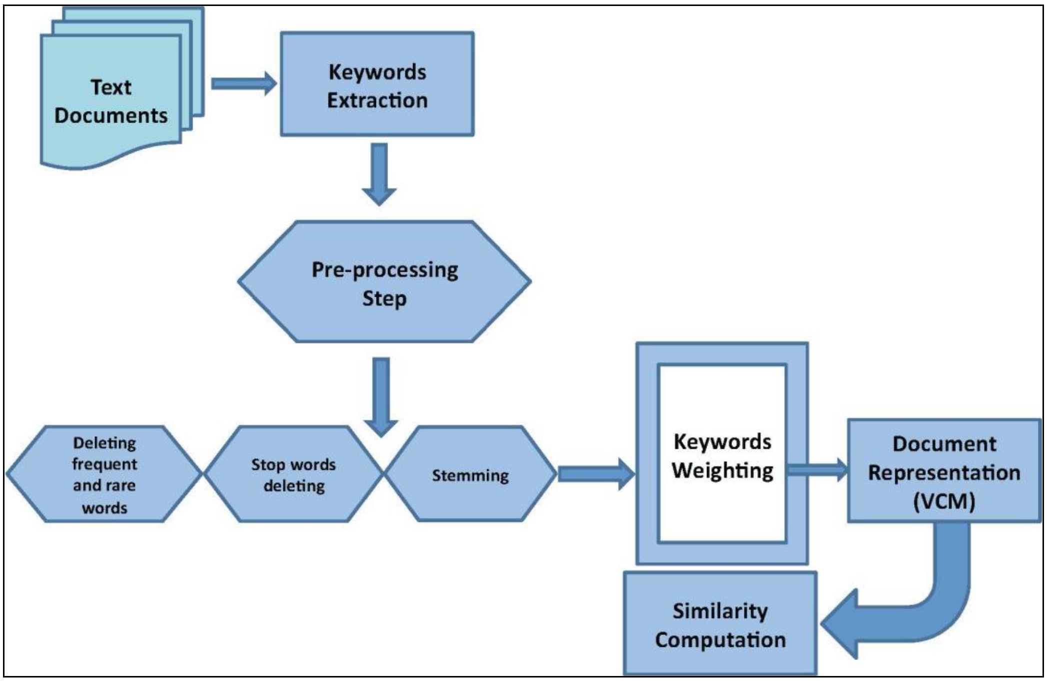

Figure 5-1: Processing Flow

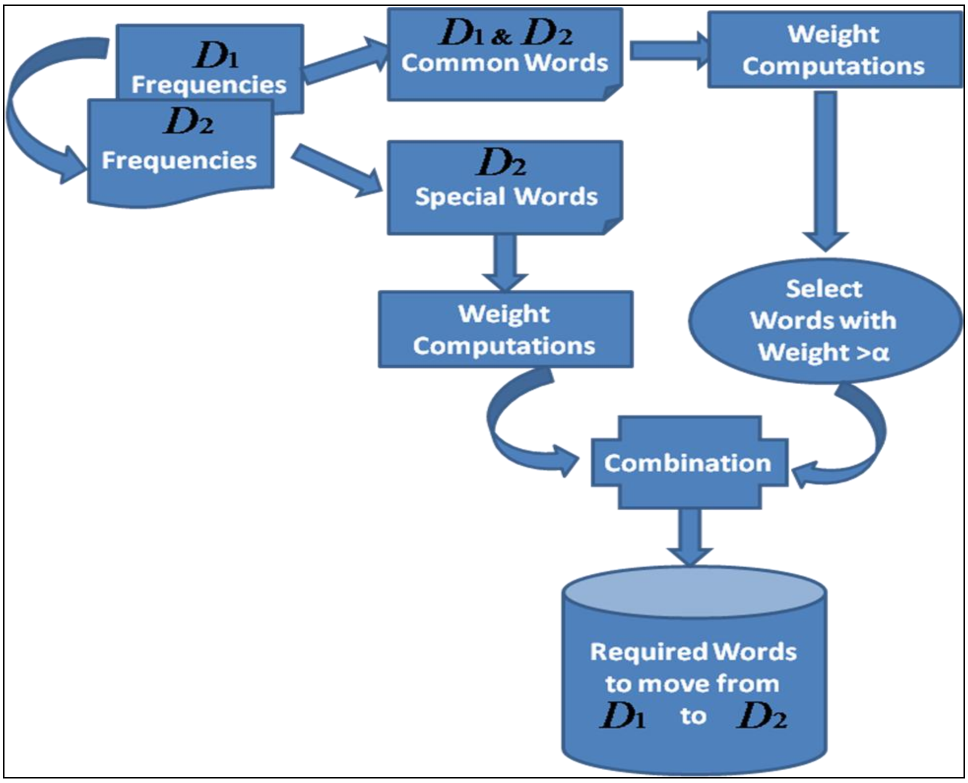

Figure 5-2: D1→ D2 transferring process

5.3 TRANSITION BETWEEN SPECIAL FIELDS

Companies and organizations operate in an increasingly competitive environment, which drives a need for continuous employee skill development. The rapid pace of technological change requires everyone to continue learning throughout life. Therefore, determining the required skills or learning for each job category or special field is very important for IT human resources. Also, clarifying what are the required skills for moving from one field to another is very important. Moreover, this study adopt a method to determine what are the important keywords required to move between different fields. The main issue in this method is the keywords of each special field.

5.3.1 Proposed Method

The goal of this method is to clarifying the important keywords to transit between different jobs. These keywords can be helpful to know the requirements to change engineers' career paths. Every two special fields have two kinds of keywords, *common keywords* (words appeared in both fields) and special keywords (words concerned to a specific field). Each kind of the keywords has its own weight formula as follow.

The weight of *special keywords* is computed by the following proposed formula:

$$S.W_k = \log(tf_{i,k}, Average(tf_{i,1}, tf_{i,2}, ..., tf_{i,n})) \qquad (5\text{-}3)$$

Where, $S.W_k$ is the weight of a special keyword k in a document i, $tf_{i,k}$ is the frequency of the keyword k in the document i, $k = 1,2,...,n$ is the keywords in a document, $i = 1,2,...,35$ is one of the 35 documents of ITSS special fields.

The weight of *common keywords* is computed by the following proposed formula:

$$C.W_{i,k} = P(tf_{i,k}) * (tf_{i,k} * IDF_i) \qquad (5\text{-}4)$$

$$C.W_{j,k} = P(tf_{j,k}) * (tf_{j,k} * IDF_j) \qquad (5\text{-}5)$$

Where, $P(tf_{i,k})$ and $P(tf_{j,k})$ is the probability of $tf_{i,k}$ and $tf_{j,k}$, respectively,

$$P(tf_{i,k}) = \frac{tf_{i,k}}{tf_{i,k} + tf_{j,k}},$$

$$P(tf_{j,k}) = \frac{tf_{j,k}}{tf_{i,k} + tf_{j,k}}, \qquad (5\text{-}6)$$

$$P(tf_{i,k}) + P(tf_{j,k}) = 1,$$

$C.W_{i,k}$, $C.W_{j,k}$ are the weight of common keyword K in a document i and a document j respectively. $tf_{i,k}, tf_{j,k}$ are the frequency of common keyword k in the documents i and j, respectively. IDF_i , IDF_j is the inverse document frequency for a document i and a document j , respectively.

So, now the two documents i and j has two kinds of keywords, Special keywords and common keywords. Every keyword has its weight. The required keywords to transfer from document i to document j are as follows:

- The *special keywords* for document *j* and,
- The *common keywords* in document *j* that have $R \geq \alpha$ .Where

  $R = \frac{C.W_{j,k}}{C.W_{i,k}}$, α is a threshold. After many trials, this study assumed that $\alpha = 5$

Figure 5-2 shows the outline of process of the transition from document D_1 to document D_2. The required words to move are the selected words with weight value greater than a threshold α and the special words of document D_2.

5.4 EXPERIMENT AND RESULTS

This experiment applied the proposed procedure on 35 text documents. Each document represents one special filed in ITSS. These documents were published by API. The first step is keywords extraction process. AnalogX keyword extraction tool to extract keywords from every document is used. The number of the keywords of all the 35 documents is 5162 keywords. This set of extracted keywords contains a lot of repeated words and a lot of stop words and so on. So, it is necessary to perform the pre-processing step. These steps contains stop words deletion and applying stemming process. The stop words were removed manually and then, stemming is applied by using Porter stemming algorithm. The number of obtained keywords after applying pre-processing step is filtered to 124 keywords.

The third step is the weight computation. The weight of keywords was computed using the formula (5-1) to represent the documents mathematically using vector space model. This process, we have 35 document vectors, each vector has 124 dimensions. Finally, the cosine similarity was computed. Table 5-1 shows the result of cosine similarity computations for 35 special fields.

Table 5-1: Cosine Similarity Result

	M1	M2	M3	S1	S2	S3	C1	C2	ITA1	ITA2	ITA3	Pm1	PM2	PM3	PM4	ITS1	ITS2	ITS3	ITS4	ITS5	ITS6	AS1	AS2	SD1	SD2	SD3	CS1	CS2	CS3	ITSM1	ITSM2	ITSM3	ITSM4	E1	E2
M1	1	0.988	0.967	0.37	0.375	0.391	0.017	0.017	0.007	0.007	0.007	0.068	0.056	0.074	0.076	0.043	0.014	0.016	0.032	0.041	0.014	0.014	0.009	0.046	0.033	0.032	0.041	0.026	0.046	0.059	0.058	0.034	0.078	0.048	0.037
M2	0.988	1	0.967	0.382	0.392	0.414	0.021	0.021	0.012	0.012	0.015	0.047	0.065	0.066	0.019	0.025	0.007	0.009	0.032	0.019	0.003	0.01	0.007	0.056	0.042	0.04	0.055	0.035	0.051	0.081	0.046	0.039	0.099	0.056	0.03
M3	0.967	0.967	1	0.336	0.345	0.36	0.011	0.011	0.009	0.009	0.009	0.068	0.061	0.073	0.075	0.017	0.006	0.008	0.026	0.016	0.003	0.003	0.003	0.053	0.04	0.038	0.036	0.022	0.089	0.108	0.036	0.024	0.061	0.04	0.028
S1	0.37	0.382	0.336	1	0.98	0.954	0.223	0.223	0.02	0.02	0.018	0.102	0.104	0.057	0.067	0.014	0.006	0.006	0.032	0.02	0.012	0.004	0.005	0.024	0.018	0.017	0.019	0.016	0.029	0.076	0.062	0.105	0.072	0.028	0.026
S2	0.375	0.392	0.345	0.98	1	0.97	0.089	0.089	0.026	0.026	0.025	0.119	0.119	0.073	0.087	0.018	0.012	0.011	0.034	0.032	0.014	0.005	0.007	0.049	0.037	0.035	0.034	0.026	0.039	0.101	0.077	0.118	0.103	0.034	0.029
S3	0.391	0.414	0.36	0.954	0.97	1	0.088	0.088	0.024	0.024	0.023	0.12	0.121	0.084	0.087	0.015	0.008	0.008	0.049	0.05	0.013	0.005	0.008	0.04	0.029	0.028	0.058	0.047	0.049	0.112	0.077	0.153	0.159	0.054	0.029
C1	0.017	0.021	0.011	0.223	0.089	0.088	1	0.999	0.008	0.008	0.008	0.045	0.042	0.035	0.039	0.008	0.004	0.003	0.004	0.015	0.008	0.009	0.013	0.006	0.003	0.002	0.005	0.003	0.013	0.03	0.027	0.025	0.026	0.07	0.012
C2	0.017	0.021	0.011	0.223	0.089	0.088	0.999	1	0.015	0.015	0.015	0.045	0.042	0.035	0.039	0.013	0.004	0.003	0.004	0.015	0.008	0.009	0.015	0.019	0.011	0.011	0.005	0.003	0.013	0.03	0.027	0.025	0.026	0.069	0.012
ITA1	0.007	0.012	0.009	0.02	0.026	0.024	0.008	0.015	1	1	0.974	0.181	0.19	0.171	0.168	0.114	0.049	0.061	0.131	0.212	0.061	0.144	0.075	0.213	0.144	0.141	0.021	0.018	0.115	0.062	0.047	0.023	0.035	0.247	0.126
ITA2	0.007	0.012	0.009	0.02	0.026	0.024	0.008	0.015	1	1	0.974	0.181	0.19	0.171	0.168	0.114	0.049	0.061	0.131	0.212	0.061	0.144	0.075	0.213	0.144	0.141	0.021	0.018	0.115	0.062	0.047	0.023	0.247	0.247	0.126
ITA3	0.007	0.015	0.009	0.018	0.025	0.023	0.008	0.015	0.974	0.974	1	0.174	0.186	0.172	0.16	0.15	0.111	0.059	0.189	0.222	0.069	0.152	0.079	0.218	0.148	0.144	0.032	0.023	0.156	0.082	0.077	0.032	0.049	0.257	0.135
Pm1	0.068	0.047	0.068	0.102	0.119	0.12	0.045	0.045	0.181	0.181	0.174	1	0.918	0.89	0.836	0.269	0.081	0.073	0.074	0.292	0.187	0.087	0.08	0.095	0.064	0.061	0.099	0.048	0.147	0.326	0.314	0.27	0.331	0.235	0.096
PM2	0.056	0.065	0.061	0.104	0.119	0.121	0.042	0.042	0.19	0.19	0.186	0.918	1	0.816	0.747	0.236	0.076	0.071	0.07	0.263	0.157	0.09	0.108	0.103	0.066	0.062	0.105	0.051	0.153	0.306	0.306	0.301	0.316	0.35	0.092
PM3	0.074	0.066	0.073	0.057	0.073	0.084	0.035	0.035	0.171	0.171	0.172	0.89	0.816	1	0.82	0.364	0.229	0.086	0.154	0.323	0.215	0.166	0.328	0.11	0.074	0.07	0.219	0.124	0.33	0.325	0.318	0.251	0.382	0.208	0.097
PM4	0.076	0.019	0.075	0.067	0.087	0.087	0.039	0.039	0.168	0.168	0.16	0.836	0.747	0.82	1	0.254	0.103	0.095	0.084	0.322	0.211	0.098	0.122	0.328	0.226	0.42	0.111	0.375	0.194	0.245	0.279	0.288	0.27	0.2	0.098
ITS1	0.043	0.025	0.017	0.014	0.018	0.015	0.008	0.013	0.114	0.114	0.15	0.269	0.236	0.364	0.254	1	0.291	0.415	0.322	0.511	0.318	0.482	0.488	0.184	0.079	0.071	0.262	0.205	0.396	0.138	0.376	0.258	0.158	0.22	0.192
ITS2	0.014	0.007	0.006	0.006	0.012	0.008	0.004	0.004	0.049	0.049	0.111	0.081	0.076	0.229	0.103	0.291	1	0.147	0.151	0.406	0.349	0.253	0.218	0.077	0.028	0.024	0.232	0.136	0.488	0.266	0.266	0.227	0.194	0.044	0.046
ITS3	0.016	0.009	0.008	0.006	0.011	0.008	0.003	0.003	0.061	0.061	0.059	0.073	0.071	0.086	0.095	0.415	0.147	1	0.129	0.353	0.16	0.203	0.142	0.06	0.03	0.027	0.065	0.121	0.144	0.04	0.253	0.223	0.052	0.061	0.112
ITS4	0.032	0.032	0.026	0.032	0.034	0.049	0.004	0.004	0.131	0.131	0.189	0.074	0.07	0.154	0.084	0.322	0.151	0.129	1	0.293	0.199	0.363	0.412	0.09	0.026	0.02	0.223	0.134	0.346	0.145	0.277	0.139	0.16	0.141	0.066
ITS5	0.041	0.019	0.016	0.02	0.032	0.05	0.015	0.015	0.212	0.212	0.222	0.292	0.263	0.323	0.322	0.511	0.406	0.353	0.293	1	0.343	0.375	0.256	0.219	0.066	0.052	0.27	0.198	0.389	0.242	0.561	0.383	0.37	0.098	0.074
ITS6	0.014	0.003	0.003	0.012	0.014	0.013	0.008	0.008	0.061	0.061	0.069	0.187	0.157	0.215	0.211	0.318	0.349	0.16	0.199	0.343	1	0.277	0.195	0.04	0.018	0.016	0.069	0.052	0.199	0.184	0.35	0.329	0.271	0.048	0.05
AS1	0.014	0.01	0.003	0.004	0.005	0.005	0.009	0.009	0.144	0.144	0.152	0.087	0.09	0.166	0.098	0.482	0.253	0.203	0.363	0.375	0.277	1	0.667	0.159	0.035	0.024	0.339	0.238	0.384	0.129	0.304	0.231	0.079	0.088	0.113
AS2	0.009	0.007	0.003	0.005	0.007	0.008	0.013	0.015	0.075	0.075	0.079	0.08	0.108	0.328	0.122	0.488	0.218	0.142	0.412	0.256	0.195	0.667	1	0.154	0.059	0.051	0.383	0.252	0.473	0.096	0.222	0.191	0.07	0.103	0.077
SD1	0.046	0.056	0.053	0.024	0.049	0.04	0.006	0.019	0.213	0.213	0.218	0.095	0.103	0.11	0.328	0.184	0.077	0.06	0.09	0.219	0.04	0.159	0.154	1	0.603	0.82	0.262	0.502	0.086	0.209	0.283	0.338	0.073	0.073	0.104
SD2	0.033	0.042	0.04	0.018	0.037	0.029	0.003	0.011	0.144	0.144	0.148	0.064	0.066	0.074	0.226	0.079	0.028	0.03	0.026	0.066	0.018	0.035	0.059	0.603	1	0.068	0.378	0.039	0.098	0.057	0.098	0.281	0.025	0.049	0.059
SD3	0.032	0.04	0.038	0.017	0.035	0.028	0.002	0.011	0.141	0.141	0.144	0.061	0.062	0.07	0.42	0.071	0.024	0.027	0.02	0.052	0.016	0.024	0.051	0.82	0.068	1	0.056	0.741	0.035	0.087	0.035	0.17	0.019	0.048	0.054
CS1	0.041	0.055	0.036	0.019	0.034	0.058	0.005	0.005	0.021	0.021	0.032	0.099	0.105	0.219	0.111	0.262	0.232	0.065	0.223	0.27	0.069	0.339	0.383	0.262	0.378	0.056	1	0.323	0.421	0.237	0.349	0.453	0.336	0.164	0.084
CS2	0.026	0.035	0.022	0.016	0.026	0.047	0.003	0.003	0.018	0.018	0.023	0.048	0.051	0.124	0.375	0.205	0.136	0.121	0.134	0.198	0.052	0.238	0.252	0.502	0.039	0.741	0.323	1	0.3	0.185	0.267	0.441	0.272	0.098	0.067
CS3	0.046	0.051	0.089	0.029	0.039	0.049	0.013	0.013	0.115	0.115	0.156	0.147	0.153	0.33	0.194	0.396	0.488	0.144	0.346	0.389	0.199	0.384	0.473	0.086	0.098	0.035	0.421	0.3	1	0.415	0.367	0.237	0.381	0.23	0.137
ITSM1	0.059	0.081	0.108	0.076	0.101	0.112	0.03	0.03	0.062	0.062	0.082	0.326	0.306	0.325	0.245	0.138	0.266	0.04	0.145	0.242	0.184	0.129	0.096	0.209	0.057	0.087	0.237	0.185	0.415	1	0.536	0.445	0.686	0.092	0.099
ITSM2	0.058	0.046	0.036	0.062	0.077	0.077	0.027	0.027	0.047	0.047	0.077	0.314	0.306	0.318	0.279	0.376	0.266	0.253	0.277	0.561	0.35	0.304	0.222	0.283	0.098	0.035	0.349	0.267	0.367	0.536	1	0.678	0.545	0.074	0.181
ITSM3	0.034	0.039	0.024	0.105	0.118	0.153	0.025	0.025	0.023	0.023	0.032	0.27	0.301	0.251	0.288	0.258	0.227	0.223	0.139	0.383	0.329	0.231	0.191	0.338	0.281	0.17	0.453	0.441	0.237	0.445	0.678	1	0.546	0.115	0.099
ITSM4	0.078	0.099	0.061	0.072	0.103	0.159	0.026	0.026	0.035	0.247	0.049	0.331	0.316	0.382	0.27	0.158	0.194	0.052	0.16	0.37	0.271	0.079	0.07	0.073	0.025	0.019	0.336	0.272	0.381	0.686	0.545	0.546	1	0.156	0.144
E1	0.048	0.056	0.04	0.028	0.034	0.054	0.07	0.069	0.247	0.247	0.257	0.235	0.35	0.208	0.2	0.22	0.044	0.061	0.141	0.098	0.048	0.088	0.103	0.073	0.049	0.048	0.164	0.098	0.23	0.092	0.074	0.115	0.156	1	0.389
E2	0.037	0.03	0.028	0.026	0.029	0.029	0.012	0.012	0.126	0.126	0.135	0.096	0.092	0.097	0.098	0.192	0.046	0.112	0.066	0.074	0.05	0.113	0.077	0.104	0.059	0.054	0.084	0.067	0.137	0.099	0.181	0.099	0.144	0.389	1

The obtained results are divided into 3 types, big values, medium values and small values.

- Big values mean that the two special fields are so similar and these values has dark gray color . These kinds of special fields have a very high chance to be in a same cluster.
- Medium values mean that the two special fields are a little bit similar and have lightly gray color. These special fields may be in a same cluster.
- Small values mean that the two special fields are so far to each other's. These special fields sure are in different clusters.

Then , the important keywords were extracted which are the words with high weight values. After keyword extraction step, these words are analyzed. There are keywords that have big weight values. These keywords are very important in the clustering process of the special fields because these keywords carry high domain-specific information content. Every special field has its important keywords, which discriminate this field. Table 5-2 shows a part of the important keywords of some special fields.

Table 5-2: High weight keywords

Special fields	***High Weight Keywords***				
M1	Market	Strategy	Analysis	Plan	Invest
C2	Consultant	Strategy	Sales	Solution	
ITA1	Architecture	Design	Model	Function	Portable
PM4	Software	Manage	Product	Monitor	
ITS1	Platform	Protocol	Security	Install	Specialist
AS2	Install	System	Web	Specialty	Interface
SD1	Software	Operation	Function	Hardware	System
CS3	Install	Network	Service	Manage	
ITSM1	Service	Operation	Contract	Service	Security
E1	Plan	Organization	Profit		

Example 5-1

This section uses a specific example shown in Figure 5-3 in order to describe the process flow of the proposed method. Two special fields Marketing Management"M1" and Sales Channel Strategy"M2" are used. First of all, the keywords were extracted from the two documents M1" and "M2" using AnalogX tool. Second, pre-processing step was conducted. Third, "M1" and "M2" were represented using vector space model. The obtained vectors are as follow:

Keywords= {accept, acquist, addit... vendor, volatil, web}

M1=(0,0,0,0,0,0,0.024,0,0,0,0,0,0,0,0.008,0.01,0,0,0,0,0,0,0.008,0.006,0,0,0.004,0,0,0,0,0,0,0,0,0,0,0.012,0,0,0,0,0,0,0,0.014,0,0,0,0,0,0.007,0.202,0,0,0,0,0,0.0101,0,0.011,0.002,0.009,0,0,0,0.014,0,0.008,0,0,0,0.013,0,0,0.014,0,0,0,0,0,0,0,0.008,0,0,0.014,0,0.014,0,0,0,0,0.011,0.002,0,0,0,0,0.01,0.013,0,0,0,0,0.007,0,0,0.004,0.036,0,0,0,0,0,0,0,0,0,0,0,0.014,0)

M2=(0,0,0,0,0,0,0.014,0,0,0,0,0,0,0,0.010,0.012,0,0,0,0,0,0,0.010,0.009,0,0,0.002,0,0,0,0,0 ,0,0,0,0,0,0.008,0,0,0,0,0,0,0,0.0169,0,0,0,0,0,0.0017,0.182,0,0,0,0,0,0,0,0.013,0,0.011,0.01,0,0,0.009,0,0.009,0,0,0,0.0198,0,0,0.016,0,0,0,0,0.005,0,0,0.01,0,0,0.016,0,0.016,0,0,0,0,0.01,0.003,0,0,0,0.013,0.012,0.019,0,0,0,0.004,0,0,0,0.004,0.04,0,0,0,0,0,0,0,0,0,0,0,0.0166,0)

Now, computing the similarity between *M1* and *M2* by using *cosine similarity*.

We obtained $Sim(M1, M2) = 0.9878$ which is very big value. This means that *M1* and *M2* are so similar; *M1* and *M2* have high chance to be in the same cluster. So, the fields *M1* and *M2* are very near to each other's. If an IT engineer wants to move from *M1* ↔ *M2*, he only needs to acquire some little skills because the similarity is very big. The required keywords to move from *M1* ↔ *M2* is shown in Figure 5-3 and Table 5-3. Formulas 2, 3 and 4 were used for computing the weight of special and common words` weight. Table 5-3 shows *M1* special words, *M2* special words and *M1& M2* common words (gray color). The required keywords to move between *M1*, *M2* is also obtained.

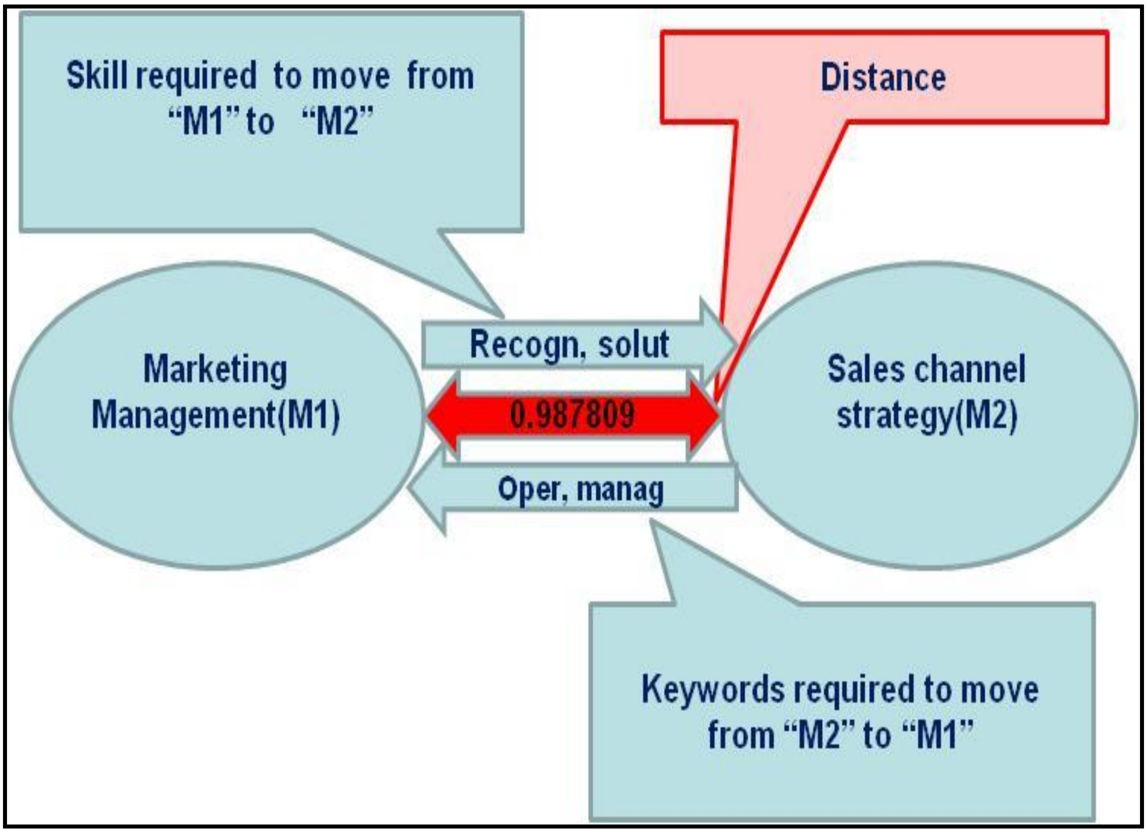

Figure 5-3: M1 <--> M2

Table 5-3: Transformation between M1 and M2

M1 S.W	***M1Wt***	***M2 S.W***	***M2 wt***	***M1&M2 C.W***	***M1 wt***	***M2 wt***	***M1->M2***	***M2->M1***
monitor	1	recogn	0	Analysis	3.768	0.942	recogn	monitor
oper	1	Optim	1.199	communicate	0.961	0.961	optim	oper
specialist	1	Segment	1.199	Competit	1.148	1.148	segment	specialist
		Solut	1.199	Cultur	0.961	0.961	solut	manag
				Custom	0.660	0.951		
				Design	0.584	0.146		
				implement	1.814	0.560		
				Invest	1.60	1.600		
				Manag	1.408	0.056		
				Market	26.26	14.77		
				novelti	1.267	1.267		
				Opportun	1.048	1.048		
				Plan	2.109	0.651		
				Polici	0.884	0.884		
				Product	1.45	2.033		
				Program	1.60	1.60		
				Releas	0.961	0.961		
				Research	1.60	1.60		
				Retent	1.60	1.60		
				Sale	1.465	0.824		
				Satisfact	0.332	0.332		
				select	1.148	1.148		
				Servic	1.456	2.033		
				Standard	0.470	0.470		
				Strategi	4.183	4.183		
				Volatil	1.60	1.60		

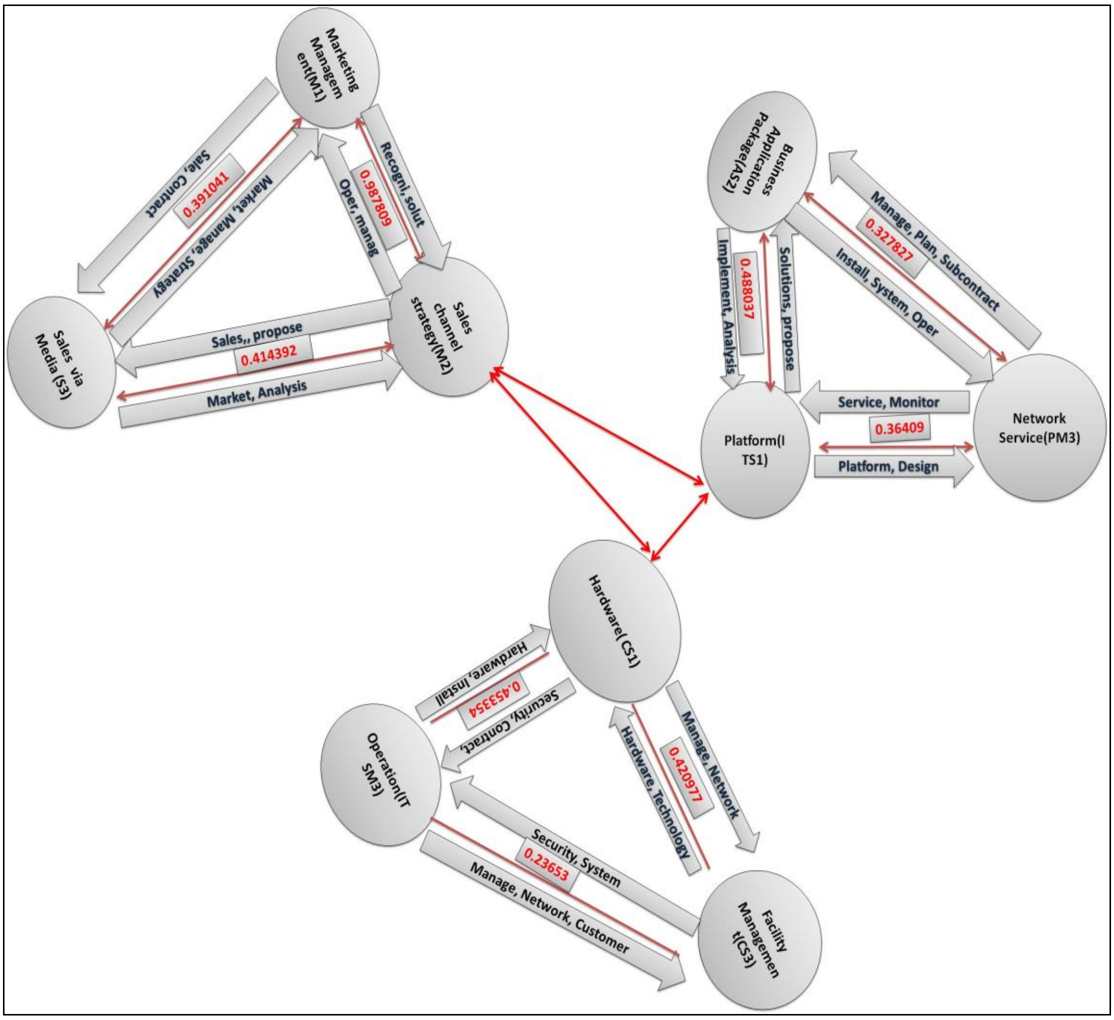

Figure 5-4: Transferring between different special fields

Figure 5-4 shows the required keywords to move between some special fields as *M1, M2, S3, PM3, AS2, ITS1, CS1, CS3, ITSM3.* The figure is composed of 3 blocks according to the similarity values. The blocks are *B1= (M1, M2, S3), B2= (PM3, AS2, ITS1),* and *B3= (CS1, CS3, ITSM3).* The distance between the special fields in the same block are corresponding to *big similarity value* and *medium similarity value* in Table 5-1. However, the distance between special fields in different blocks correspond to *small similarity values* in Table 5-1.

Figure 5-4 shows that the distance between *B2* and *B3* is smaller than the distance between *(B1<-->B2)* and *(B1<-->B3).* This means that the similarity between *(B2<-->B3)* is bigger than the similarity between *(B1<-->B2)* and *(B1<-->B3).* This study investigated that, if an engineer is in a specific job, he can move to another job in the same block by a little training and studying. But, if he moves to another job in a different block, he needs a lot of training and study to acquire the required skills for the new job.

5.5 K-MEANS AND ITSS SPECIAL FIELDS

K-means is one of the simplest unsupervised learning algorithms to group similar data objects. This algorithm was developed by J. MacQueen (1967) and then by J.A. Hartigan and M.A. Wong around 1975 [Krishna, et al., 2010; Jain, et al., 1988; Xu, et al., 2005]. K-means forms clusters for n objects based on the attributes into k partitions where $k<n$. The algorithm starts by partitioning the input points into k initial sets, either at random or using heuristic data. Then calculates the mean point or centroid of each set. A new partition is constructed by associating each point with the closest centroid. Then the centroids are recalculated for new clusters, and the algorithm is repeated by alternate application of these two steps until convergence, which is obtained when the points no longer switch clusters. The centroids should be placed in a cunning way as different centroid location provides different results. The algorithm converges when there is no further change in assignment of instances to clusters.

We applied K-means algorithm on 35 special fields' documents with 4 clusters. We initially supposed those clusters are *M1, PM3, ITS1*, and *E1*. After 2 iterations we obtained 4 clusters as in Figure 5-5.

Obtained results of k-means and the result of the proposed method, which derived the required keywords to move from one special field to another, are analyzed. By comparing k-means algorithm result and the result in Figure 5-6, we discovered that the special fields in the same cluster require a small number of keywords to move between clusters. While, the special fields in different clusters require a big number of keywords to move between each other’s (between clusters). For example:

- From $M1 \rightarrow M2$ are 5 keywords.
- From $S1 \rightarrow S2$ are 2 keywords.
- From $S2 \rightarrow S1$ are non-keywords.
- From $M1 \rightarrow ITA1$ are 33keywords.
- From $ITS2 \rightarrow M1$ are 24 keywords..

So, we can say that the proposed method to extract the keywords required to move between fields is good, because the result of k-means algorithm has a same meaning.

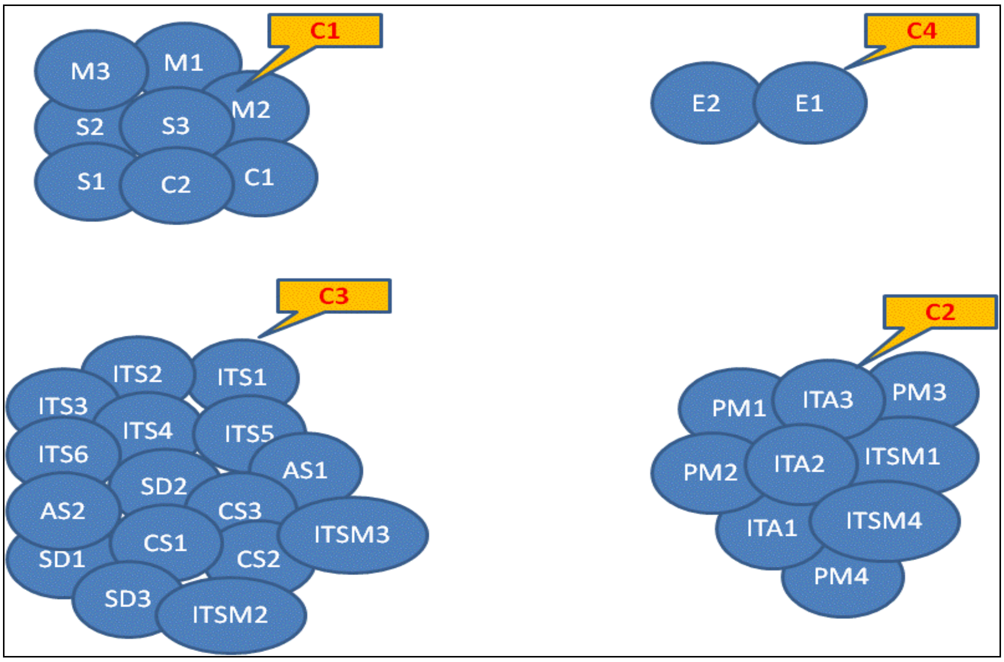

Figure 5-5: K-Means resulted clusters

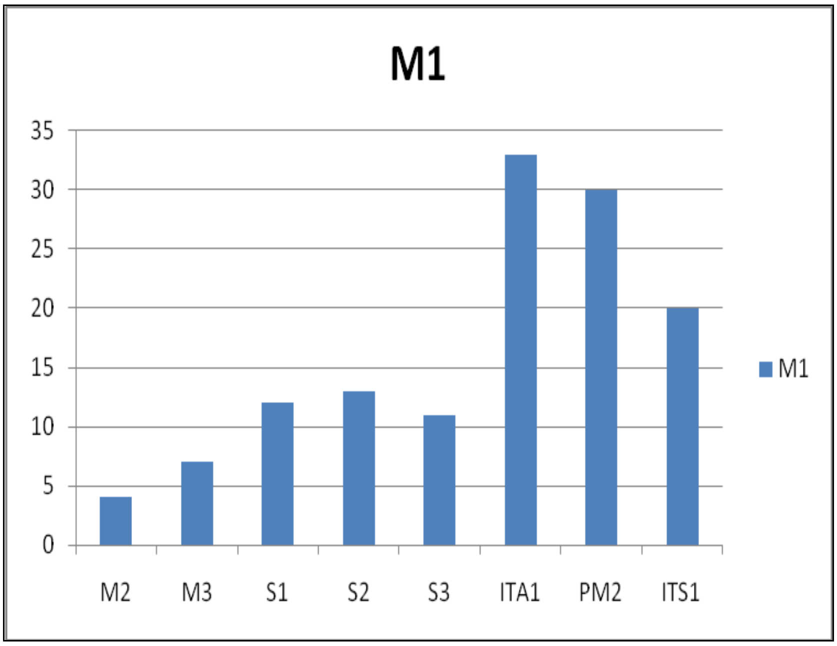

Figure 5-6: Number of keywords required to move from M1 to other fields

5.6 DISCUSSIONS

In order to verify the achievement of this study, 6 experts in software are asked to take a discussion. Information about these six experts are shown in the Table 5-4. The term to evaluate is keywords. The keywords show the terms which need to learn for transferring specialty fields shown in Figure 5-4 on the distance between specialty fields are shown in Table 5-1. As the distance between specialty fields shown in Figure 5-4 is short in every job category, the experts highly evaluate that the computation results is regarded as the reliable results.

Moreover, as the distance between specialty fields belonging to "Marketing" and "Sale" is short and the distance from another specialty fields is long, the experts highly evaluate that reliability of the computation results is improved. In addition, the experts evaluate that the results are also unpredictable because the distance among six specialty fields belonging to "IT Specialist" and two specialty fields belonging to "Application Specialist" is short. The reason is those specialty fields consist of independent technical terms required for software development such as "IT Specialist"= {Platform, Network, Database, Common Application Infrastructure, System Management, Security}.

Table 5-4: Experts Information

	Professional affiliation	***Specialty***	***Work experience***
1	Felicanetworks	Software Test	10~15
2	Mitsubishi Electric Co. Ltd.	Software Business management	20~25
3	Hitachi, Ltd	Software development	15~20
4	NTT DATA Ltd.	Software production control	20~25
5	TIS Inc	Software skill person education	20~25
6	Japan Users Association of Information Systems	Software Lecturer	25~30

Therefore, most of the experts estimate that the distance among 6 specialty fields belonging to "IT Specialist" is long unlike in the case of other job categories. However, the experts highly evaluate that the results are new discoveries because the results are computed correctly with reconfirming the required skills for six specialty fields in the ITSS document. Moreover, the experts highly evaluate that other new discoveries are the distance among "Application Specialist"= {Application System, Application Package} and six specialty fields belonging to "IT Specialist" is short.

The experts are unpredictable that the distance among "Education"= {Training Planning, Instructions} and ten other job categories is long because "Education" requires for both a skill to teach and a skill of technical terms to teach. But the experts have opinions about the necessity of improvements in the ITSS document because experts found out that "Education" in the ITSS document only focus on the skill to teach. Consequently, the experts highly evaluate to obtain new knowledge and the improvement plans from the achievement of this study. The experts generally agree with skills required when engineers transfer from the present specialty field to other specialty fields as shown in Figure 5-4. Furthermore, as the distance between specialty fields is regarded as an image, the experts highly evaluate that the results clearly show engineers the guidelines to improve their skills in the future.

CHAPTER-6

Comparison of ITSS Definition and A Questionnaire to Improve Software engineer's Skill

6.1 INTRODUCTION

This chapter compares between the obtained results in chapter 5 and the results of a questionnaire that was applied on a Japanese company uses ITSS to develop its human resources. The purpose of this questionnaire is to grasp the skills of the employees . In this comparison, we tried to process the data of the questionnaire to be compared with the obtained results in chapter 5. Section 6.2 describes the questionnaire. Section 6.3 explains the processing of the questionnaire's data. Finally, section 6.3.2 computed the correlation coefficient between the results of the questionnaire and the obtained results on chapter 5 .

6.2 QUESTIONNAIRE

This section describes a questionnaire that was created by Chikako Morimoto and applied to a Japanese company called " system integration business" to grasp the skills of employees . The questionnaire was directed to 1080 of the employees in a financial services part of the company. Ms. Morimoto asked the employees some questions related to the different special fields. These questions are about the skill area of each ITSS's special fields. Table 6-1 shows the summary of the questionnaire.

Figure 6-1 is a snapshot of the questionnaire. In Figure 6-1, Job category column is the 9 ITSS's job categories, Special Fields are some of ITSS's special fields, Post is some departments in the company and finally, the values in these figures are the answer of the questions. If there is no value, this means that this employee does not have this skill. The values ranged from 0 to 7.

Table 6-1: Questionnaire Summary

Investigation company	System integration business (The candidate for investigation is a financial-services part)
Investigation purpose	For an employee's skill grasp
Number for investigation	1080
Reply number	1014
Response rate	93.9％
Consultation period	February 13, 2012 to March 23, 2012
Investigation item	Skill investigation based on ITSS The experienced operating field The experienced development process Command of English 1. is used for analysis among the above.
Examination method	The self-reply by the questionnaire system using in-company intranet

Job Category	Specialty Field		Question No	No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	:
		コンサルティングメソドロジの活用	71						4			6												
		解決の方向性の提言	70						4			6												
		「経営におけるIT」の知見からの課題発見	69						4			6												
		ビジネス上の課題の特定と分析	68						4			6												
Sales	Sales via Media	セールスメディアの活用	65			4			4			5	5					4						
		ネゴシエーション	64			4			4		3	5	5					4					3	
		コミュニケーション	63			4			4		4	5	5					4					3	
		リーダシップ	62			4			4		3	5	5					4					3	
		プロジェクトマネジメント	61			4			4			4	5					4						
		ビジネス戦略	60			4			4		3	5	5					4					3	
		セールス事務管理	59			4			4		4	4	5					4					3	
		顧客満足度管理	58			4			4		4	5	5					4					3	
		ITソリューション提案	57			4			4		4	5	5					4					3	
		顧客環境分析	56			4			4		3	5	5					4					3	
	Product Sales by Visiting Customers	特定製品、サービステクノロジ	54			5	4		4		4	6	5	4				5		5	4	3	3	
		ネゴシエーション	53			5	4		4		3	6	5	5				5		5	5	3	3	
		コミュニケーション	52			5	4		4		4	6	5	5				5		5	5	3	3	
		リーダシップ	51			5	4		4		4	6	5	5				5		5	5	3	3	
		プロジェクトマネジメント	50			5	4		4			5	5	5				5		5	5			
		ビジネス戦略	49			5	4		4		3	5	5	5				5		5	5	3	3	
		セールス事務管理	48			5	4		4		5	5	5	5				5		5	5	3	3	3
		顧客満足度管理	47			5	4		4		4	6	5	5				5		5	5	3	3	
		ITソリューション提案	46			5	4		4		4	6	5	5				5		5	4	3	3	
		顧客環境分析	45			5	4		4		4	6	5	5				5		5	4	3	3	
	Consulting Sales by Visiting Customers	顧客リレーションシップ構築	43			6	5	3	4		5	6	6	6		3		5	6	6	6		3	
		ネゴシエーション	42		6	6	5	3	4		4	7	6	4		3		5	6	6	6		3	
		コミュニケーション	41		6	6	5	3	4		4	7	6	4		3		5	6	6	6		3	
		リーダシップ	40		6	6	5	3	4		4	7	6	4				5	6	6	6			
		プロジェクトマネジメント	39			6	5	3	4			6	6			3		5	6	6	6			
		ビジネス戦略	38		6	6	5	3	3		4	6	6	4		3		5	6	6	5		3	
		セールス事務管理	37		6	6	5	3	3		6	6	6	4		3		5	6	6	6		3	3
		顧客満足度管理	36			6	5	3	3		4	6	6			3		5	6	6	5		3	
		ITソリューション提案	35		6	6	5	3	3		4	7	6	4		3		5	6	6	5		3	
		顧客環境分析	34			6	5	3	5		4	6	6	4		3		5	6	6	5			
Marketing	Market Communication	マーケットコミュニケーション戦略	31			6			4		3	6	5							5				
		ネゴシエーション	30			6			3		3	6	5							5			3	
		コミュニケーション	29			6			3		3	6	5							5			3	
		リーダシップ	28			6	4		4		3	6	5							5				
		プロジェクトマネジメント	27			6	4		3			6	5							5				
		コンプライアンス	26			6	4		3		3	6	5							5				
		マーケティング環境分析	25			6	4		3		3	6	5							5			3	
		マーケティング戦略	24			6	4		3		3	6	5							5			3	
		市場機会の評価と選定	23			6	4		3		3	6	5							5			3	
	Sales Channel Strategy	販売チャネル戦略	21			6			4			6	5							5				
		ネゴシエーション	20			6	4		4			6	5	4						5				
		コミュニケーション	19			6	4		4		4	6	5	4						5				
		リーダシップ	18			6	4		4			6	5	4						5				
		プロジェクトマネジメント	17			6	4					6	5							5				
		コンプライアンス	16			6	4		4			6	5	4						5				
		マーケティング環境分析	15			6	4		4			6	5							5				
		マーケティング戦略	14			5	4		4			6	5	4						5				
		市場機会の評価と選定	13			5	4		4			6	5	4						5				
	Marketing Management	マーケティング統括	11			6						6	5							6				
		ネゴシエーション	10			6						7	5							6				
		コミュニケーション	9			6						7	5							5				
		リーダシップ	8			6						7	5							6				
		プロジェクトマネジメント	7			5						6	5							5				
		コンプライアンス	6			6						6	5							6				
		マーケティング環境分析	5			6						6	5							5				
		マーケティング戦略	4			6						6	5							5				
		市場機会の評価と選定	3			6			5			6	5							6				
				post	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	

Figure 6-1: Snapshot of the Questionnaire

6.3 DOCUMENT ANALYSIS AND QUESTIONNAIRE

This section compares between the results of the document analysis that were obtained in chapter 5 and results of the questionnaire. As mentioned in section 5.2 the cosine similarity between the 35 documents of ITSS was computed in chapter 5. Table 6-2 shows the result of cosine similarity between the service's specialty fields. In the remaining of this chapter, we converted the data of the questionnaire to a format that can be used to compute the cosine similarity between the special fields using questionnaire's data. We proposed a method to convert the questionnaire's data to the new format. This method will be explained in the next section.

Table 6-2: Cosine Similarity

	M1	M2	M3	S1	S2	S3	C1	C2	ITA1	ITA2	ITA3	Pm1	PM2	PM3	PM4	ITS1	ITS2	ITS3	ITS4	ITS5	ITS6	AS1	AS2	SD1	SD2	SD3	ITSM1	ITSM2	ITSM3	ITSM4
M1	1	0.988	0.967	0.37	0.375	0.391	0.017	0.017	0.007	0.007	0.007	0.068	0.056	0.074	0.076	0.043	0.014	0.016	0.032	0.041	0.014	0.014	0.009	0.046	0.033	0.032	0.059	0.058	0.034	0.078
M2	0.988	1	0.967	0.382	0.392	0.414	0.021	0.021	0.012	0.012	0.015	0.047	0.065	0.066	0.019	0.025	0.007	0.009	0.032	0.019	0.003	0.01	0.007	0.056	0.042	0.04	0.081	0.046	0.039	0.099
M3	0.967	0.967	1	0.336	0.345	0.36	0.011	0.011	0.009	0.009	0.009	0.068	0.061	0.073	0.075	0.017	0.006	0.008	0.026	0.016	0.003	0.003	0.003	0.053	0.04	0.038	0.108	0.036	0.024	0.061
S1	0.37	0.382	0.336	1	0.98	0.954	0.223	0.223	0.02	0.02	0.018	0.102	0.104	0.057	0.067	0.014	0.006	0.006	0.032	0.02	0.012	0.004	0.005	0.024	0.018	0.017	0.076	0.062	0.105	0.072
S2	0.375	0.392	0.345	0.98	0.98	0.97	0.089	0.089	0.026	0.026	0.025	0.119	0.119	0.073	0.087	0.018	0.012	0.011	0.034	0.032	0.014	0.005	0.007	0.049	0.037	0.035	0.101	0.077	0.118	0.103
S3	0.391	0.414	0.36	0.954	0.98	1	0.088	0.088	0.024	0.024	0.023	0.12	0.121	0.084	0.087	0.015	0.008	0.008	0.049	0.05	0.013	0.005	0.008	0.04	0.029	0.028	0.112	0.077	0.153	0.159
C1	0.017	0.021	0.011	0.223	0.98	0.088	1	0.999	0.008	0.008	0.008	0.045	0.042	0.035	0.039	0.008	0.004	0.003	0.004	0.015	0.008	0.009	0.013	0.006	0.003	0.002	0.03	0.027	0.025	0.026
C2	0.017	0.021	0.011	0.223	0.98	0.088	0.999	1	0.015	0.015	0.015	0.045	0.042	0.035	0.039	0.013	0.004	0.003	0.004	0.015	0.008	0.009	0.015	0.019	0.011	0.011	0.03	0.027	0.025	0.026
ITA1	0.007	0.012	0.009	0.02	0.98	0.024	0.008	0.015	1	1	0.974	0.181	0.19	0.171	0.168	0.114	0.049	0.061	0.131	0.212	0.061	0.144	0.075	0.213	0.144	0.141	0.062	0.047	0.023	0.035
ITA2	0.007	0.012	0.009	0.02	0.98	0.024	0.008	0.015	1	1	0.974	0.181	0.19	0.171	0.168	0.114	0.049	0.061	0.131	0.212	0.061	0.144	0.075	0.213	0.144	0.141	0.062	0.047	0.023	0.247
ITA3	0.007	0.015	0.009	0.018	0.98	0.023	0.008	0.015	0.974	0.974	1	0.174	0.186	0.172	0.16	0.15	0.111	0.059	0.189	0.222	0.069	0.152	0.079	0.218	0.148	0.144	0.082	0.077	0.032	0.049
Pm1	0.068	0.047	0.068	0.102	0.98	0.12	0.045	0.045	0.181	0.181	0.174	1	0.918	0.89	0.836	0.269	0.081	0.073	0.074	0.292	0.187	0.087	0.08	0.095	0.064	0.061	0.326	0.314	0.27	0.331
PM2	0.056	0.065	0.061	0.104	0.98	0.121	0.042	0.042	0.19	0.19	0.186	0.918	1	0.816	0.747	0.236	0.076	0.071	0.07	0.263	0.157	0.09	0.108	0.103	0.066	0.062	0.306	0.306	0.301	0.316
PM3	0.074	0.066	0.073	0.057	0.98	0.084	0.035	0.035	0.171	0.171	0.172	0.89	0.816	1	0.82	0.364	0.229	0.086	0.154	0.323	0.215	0.166	0.328	0.11	0.074	0.07	0.325	0.318	0.251	0.382
PM4	0.076	0.019	0.075	0.067	0.98	0.087	0.039	0.039	0.168	0.168	0.16	0.836	0.747	0.82	1	0.254	0.103	0.095	0.084	0.322	0.211	0.098	0.122	0.328	0.226	0.42	0.245	0.279	0.288	0.27
ITS1	0.043	0.025	0.017	0.014	0.98	0.015	0.008	0.013	0.114	0.114	0.15	0.269	0.236	0.364	0.254	1	0.291	0.415	0.322	0.511	0.318	0.482	0.488	0.184	0.079	0.071	0.138	0.376	0.258	0.158
ITS2	0.014	0.007	0.006	0.006	0.98	0.008	0.004	0.004	0.049	0.049	0.111	0.081	0.076	0.229	0.103	0.291	1	0.147	0.151	0.406	0.349	0.253	0.218	0.077	0.028	0.024	0.266	0.266	0.227	0.194
ITS3	0.016	0.009	0.008	0.006	0.98	0.008	0.003	0.003	0.061	0.061	0.059	0.073	0.071	0.086	0.095	0.415	0.147	1	0.129	0.353	0.16	0.203	0.142	0.06	0.03	0.027	0.04	0.253	0.223	0.052
ITS4	0.032	0.032	0.026	0.032	0.98	0.049	0.004	0.004	0.131	0.131	0.189	0.074	0.07	0.154	0.084	0.322	0.151	0.129	1	0.293	0.199	0.363	0.412	0.09	0.026	0.02	0.145	0.277	0.139	0.16
ITS5	0.041	0.019	0.016	0.02	0.98	0.05	0.015	0.015	0.212	0.212	0.222	0.292	0.263	0.323	0.322	0.511	0.406	0.353	0.293	1	0.343	0.375	0.256	0.219	0.066	0.052	0.242	0.561	0.383	0.37
ITS6	0.014	0.003	0.003	0.012	0.98	0.013	0.008	0.008	0.061	0.061	0.069	0.187	0.157	0.215	0.211	0.318	0.349	0.16	0.199	0.343	1	0.277	0.195	0.04	0.018	0.016	0.184	0.35	0.329	0.271
AS1	0.014	0.01	0.003	0.004	0.98	0.005	0.009	0.009	0.144	0.144	0.152	0.087	0.09	0.166	0.098	0.482	0.253	0.203	0.363	0.375	0.277	1	0.667	0.159	0.035	0.024	0.129	0.304	0.231	0.079
AS2	0.009	0.007	0.003	0.005	0.98	0.008	0.013	0.015	0.075	0.075	0.079	0.08	0.108	0.328	0.122	0.488	0.218	0.142	0.412	0.256	0.195	0.667	1	0.154	0.059	0.051	0.096	0.222	0.191	0.07
SD1	0.046	0.056	0.053	0.024	0.98	0.04	0.006	0.019	0.213	0.213	0.218	0.095	0.103	0.11	0.328	0.184	0.077	0.06	0.09	0.219	0.04	0.159	0.154	1	0.603	0.82	0.209	0.283	0.338	0.073
SD2	0.033	0.042	0.04	0.018	0.98	0.029	0.003	0.011	0.144	0.144	0.148	0.064	0.066	0.074	0.226	0.079	0.028	0.03	0.026	0.066	0.018	0.035	0.059	0.603	1	0.068	0.057	0.098	0.281	0.025
SD3	0.032	0.04	0.038	0.017	0.98	0.028	0.002	0.011	0.141	0.141	0.144	0.061	0.062	0.07	0.42	0.071	0.024	0.027	0.02	0.052	0.016	0.024	0.051	0.82	0.068	1	0.087	0.035	0.17	0.019
ITSM1	0.059	0.081	0.108	0.076	0.98	0.112	0.03	0.03	0.062	0.062	0.082	0.326	0.306	0.325	0.245	0.138	0.266	0.04	0.145	0.242	0.184	0.129	0.096	0.209	0.057	0.087	1	0.536	0.445	0.686
ITSM2	0.058	0.046	0.036	0.062	0.98	0.077	0.027	0.027	0.047	0.047	0.077	0.314	0.306	0.318	0.279	0.376	0.266	0.253	0.277	0.561	0.35	0.304	0.222	0.283	0.098	0.035	0.536	1	0.678	0.545
ITSM3	0.034	0.039	0.024	0.105	0.98	0.153	0.025	0.025	0.023	0.023	0.032	0.27	0.301	0.251	0.288	0.258	0.227	0.223	0.139	0.383	0.329	0.231	0.191	0.338	0.281	0.17	0.445	0.678	1	0.546
ITSM4	0.078	0.099	0.061	0.072	0.98	0.159	0.026	0.026	0.035	0.247	0.049	0.331	0.316	0.382	0.27	0.158	0.194	0.052	0.16	0.37	0.271	0.079	0.07	0.073	0.025	0.019	0.686	0.545	0.546	1

6.3.1 Questionnaire's data processing

The raw data of the questionnaire needs a pre-processing step to be compared with the results of the document analysis .The preprocessing step is used to reduce the complexity of the questionnaire and to be to handle. In this step, the questionnaire will be transformed from a version in which every employee has many values(answers of the questions) for each special field to a version in which every employee has only one value that represents his answers for all the questions. In the questionnaire every special field has some questions which are related to the skills required for that special field. Every employee answers each question by a value. This value ranges from 0 to 7 where zero means no answer and other values express how fit the question with the answer. The pre-processing step will be done as follows: For every special field do:

1. **Attaching only one value to every employee**

According to the questionnaire, every employee has an answer for every question(values bounded by an oval in Figure 6-2). I summed the values of these answers. So, every employee become has only one value(As example, the value 53 in Figure 6-2). Now, every special field has 1014 values, one value for each employee. We supposed that every special field is a text document and that document is represented by its keywords. The frequency of the keywords are the computed values of each employee. Figure 6-2 explains the pre-processing step.

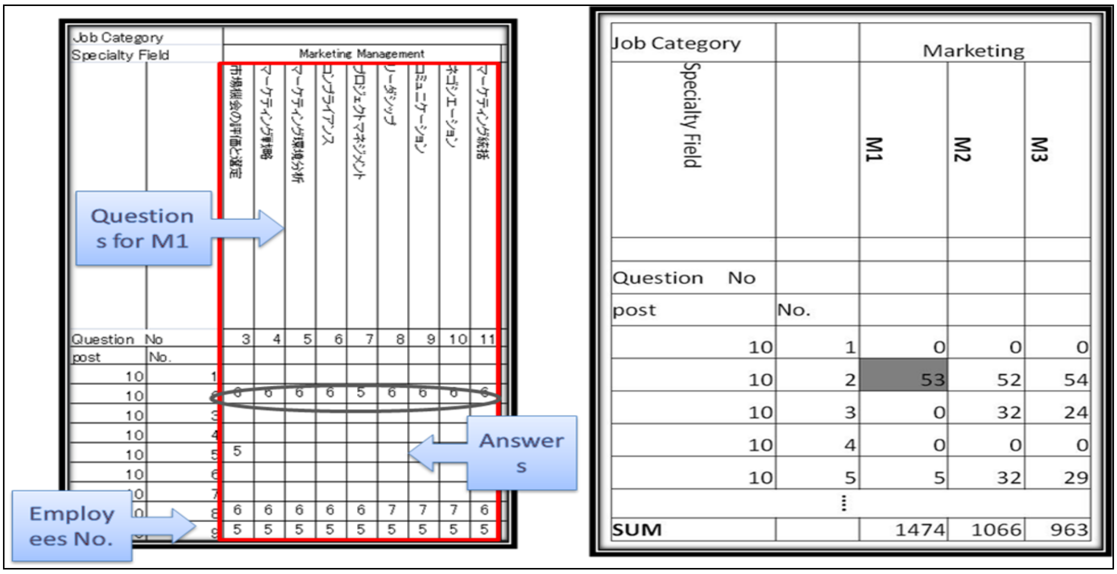

Figure 6-2: Part of the Questionnaire

2. **Representing every special field as a vector**

We supposed that every special field is a text document that can be represented by 1014 terms(These terms are the employees). Every term (employee) has a frequency(that was computed in step 1). Throughout this section, we will use the symbols S , w , n to denote a special field, the weight of terms in S , and the number of terms in S respectively.

$$S = (w_1, w_2, ..., w_n) \qquad (6\text{-}1)$$

Where w_i is the weight of the i^{th} term in the special field S , n is the number of employees.

3. Computing the weight of every employee

Compute the weight of every term(employee). The weight of a term t_i is shown in equation 6-2.

$$w_i = \frac{t_i}{\sum_{i=1}^{n} t_i} \qquad (6\text{-}2)$$

Where, t_i is the value of the answer of the i^{th} employee in the special field S .

4. Similarity Computation.

Finally, compute the cosine similarity between the 30 specialty fields by equation 6-3.

$$sim(S_i, S_j) = \frac{S_i * S_j}{\|S_i\| \times \|S_j\|} \qquad (6\text{-}3)$$

Where '*' denotes the dot-product of the two vectors S_i and S_j and $\|S_i\|$ is the length of the vector S_i . Table 6-3 shows the result of cosine similarity.

Now, we have two similarity tables. one for the obtained data in chapter 5 and the other is for the questionnaire. So, correlation coefficient was used to discuss the relation between the values of the two tables and also to evaluate the obtained results.

Table 6-3: Cosine Similarity of the Questionnaire

	M1	M2	M3	S1	S2	S3	C1	C2	ITA1	ITA2	ITA3	PM1	PM2	PM3	PM4	ITS1	ITS2	ITS3	ITS4	ITS5	ITS6	AS1	AS2	SD1	SD2	SD3	ITSM1	ITSM2	ITSM3	ITSM4
M1	1	0.778	0.752	0.51	0.427	0.381	0.577	0.486	0.002	0.002	0.002	0.001	0.023	0.006	0.002	0.182	0.116	0.162	0.182	0.176	0.179	0.232	0.274	0.084	0.104	0.104	0.063	0.073	0.052	0.058
M2	0.778	1	0.744	0.537	0.558	0.515	0.607	0.507	0.254	0.26	0.26	0.228	0.266	0.226	0.229	0.154	0.135	0.14	0.15	0.144	0.157	0.19	0.275	0.149	0.151	0.202	0.09	0.103	0.118	0.126
M3	0.752	0.744	1	0.451	0.531	0.609	0.454	0.442	0.249	0.254	0.25	0.227	0.221	0.217	0.269	0.156	0.124	0.145	0.159	0.148	0.158	0.197	0.216	0.142	0.145	0.142	0.08	0.084	0.088	0.088
S1	0.51	0.537	0.451	1	0.713	0.035	0.017	0.016	0.001	0.001	0.001	7E-04	0.018	0.005	9E-04	0.131	0.105	0.124	0.127	0.133	0.138	0.152	0.169	0.098	0.102	0.108	0.097	0.118	0.075	0.062
S2	0.427	0.558	0.531	0.713	1	0.665	0.253	0.186	0.157	0.16	0.157	0.185	0.18	0.158	0.177	0.122	0.12	0.118	0.12	0.12	0.131	0.137	0.197	0.096	0.093	0.147	0.079	0.091	0.144	0.146
S3	0.381	0.515	0.609	0.035	0.665	1	0.221	0.25	0.139	0.142	0.142	0.108	0.181	0.205	0.159	0.126	0.155	0.118	0.12	0.118	0.123	0.091	0.152	0.145	0.144	0.141	0.102	0.115	0.116	0.131
C1	0.577	0.607	0.454	0.017	0.253	0.221	1	0.821	0.285	0.289	0.284	0.287	0.279	0.163	0.209	0.135	0.107	0.147	0.139	0.129	0.142	0.248	0.258	0.102	0.091	0.231	0.066	0.067	0.11	0.119
C2	0.486	0.507	0.442	0.016	0.186	0.25	0.821	1	0.308	0.312	0.306	0.276	0.322	0.191	0.218	0.16	0.112	0.173	0.17	0.152	0.165	0.239	0.264	0.112	0.099	0.229	0.082	0.083	0.065	0.071
ITA1	0.002	0.254	0.249	0.001	0.157	0.139	0.285	0.308	1	0.998	0.996	0.563	0.138	0.39	0.457	0.597	0.367	0.585	0.582	0.582	0.589	0.508	0.387	0.338	0.285	0.302	0.318	0.327	0.196	0.136
ITA2	0.002	0.26	0.254	0.001	0.16	0.142	0.289	0.312	0.998	1	0.998	0.562	0.143	0.399	0.457	0.6	0.374	0.584	0.582	0.584	0.59	0.499	0.386	0.337	0.288	0.303	0.325	0.335	0.198	0.139
ITA3	0.002	0.26	0.25	0.001	0.157	0.142	0.284	0.306	0.996	0.998	1	0.558	0.144	0.41	0.451	0.608	0.388	0.587	0.585	0.591	0.596	0.491	0.377	0.333	0.284	0.297	0.338	0.35	0.204	0.139
PM1	0.001	0.228	0.227	7E-04	0.185	0.108	0.287	0.276	0.563	0.562	0.558	1	0.224	0.293	0.618	0.504	0.225	0.505	0.505	0.508	0.507	0.727	0.421	0.267	0.207	0.249	0.305	0.272	0.193	0.162
PM2	0.023	0.266	0.221	0.018	0.18	0.181	0.279	0.322	0.138	0.143	0.144	0.224	1	0.269	0.185	0.123	0.131	0.109	0.12	0.133	0.129	0.158	0.234	0.079	0.094	0.148	0.134	0.102	0.082	0.05
PM3	0.006	0.226	0.217	0.005	0.158	0.205	0.163	0.191	0.39	0.399	0.41	0.293	0.269	1	0.295	0.375	0.545	0.317	0.32	0.364	0.36	0.165	0.218	0.196	0.227	0.239	0.349	0.389	0.221	0.175
PM4	0.002	0.229	0.269	9E-04	0.177	0.159	0.209	0.218	0.457	0.457	0.451	0.618	0.185	0.295	1	0.439	0.218	0.456	0.47	0.443	0.453	0.532	0.463	0.318	0.232	0.314	0.206	0.197	0.173	0.148
ITS1	0.182	0.154	0.156	0.131	0.122	0.126	0.135	0.16	0.597	0.6	0.608	0.504	0.123	0.375	0.439	1	0.579	0.965	0.972	0.978	0.981	0.544	0.378	0.454	0.417	0.344	0.242	0.462	0.34	0.242
ITS2	0.116	0.135	0.124	0.105	0.12	0.155	0.107	0.112	0.367	0.374	0.388	0.225	0.131	0.545	0.218	0.579	1	0.506	0.507	0.56	0.572	0.209	0.228	0.333	0.332	0.236	0.41	0.444	0.314	0.222
ITS3	0.162	0.14	0.145	0.124	0.118	0.118	0.147	0.173	0.585	0.584	0.587	0.505	0.109	0.317	0.456	0.965	0.506	1	0.983	0.965	0.977	0.588	0.383	0.456	0.405	0.355	0.396	0.389	0.31	0.224
ITS4	0.182	0.15	0.159	0.127	0.12	0.12	0.139	0.17	0.582	0.582	0.585	0.505	0.12	0.32	0.47	0.972	0.507	0.983	1	0.971	0.981	0.586	0.392	0.471	0.422	0.364	0.405	0.39	0.312	0.222
ITS5	0.176	0.144	0.148	0.133	0.12	0.118	0.129	0.152	0.582	0.584	0.591	0.508	0.133	0.364	0.443	0.978	0.56	0.965	0.971	1	0.982	0.55	0.379	0.446	0.409	0.342	0.487	0.478	0.361	0.255
ITS6	0.179	0.157	0.158	0.138	0.131	0.123	0.142	0.165	0.589	0.59	0.596	0.507	0.129	0.36	0.453	0.981	0.572	0.977	0.981	0.982	1	0.565	0.383	0.457	0.417	0.359	0.439	0.43	0.335	0.25
AS1	0.232	0.19	0.197	0.152	0.137	0.091	0.248	0.239	0.508	0.499	0.491	0.727	0.158	0.165	0.532	0.544	0.209	0.588	0.586	0.55	0.565	1	0.569	0.349	0.288	0.336	0.2	0.171	0.168	0.164
AS2	0.274	0.275	0.216	0.169	0.197	0.152	0.258	0.264	0.387	0.386	0.377	0.421	0.234	0.218	0.463	0.378	0.228	0.383	0.392	0.379	0.383	0.569	1	0.307	0.3	0.395	0.152	0.159	0.162	0.156
SD1	0.084	0.149	0.142	0.098	0.096	0.145	0.102	0.112	0.338	0.337	0.333	0.267	0.079	0.196	0.318	0.454	0.333	0.456	0.471	0.446	0.457	0.349	0.307	1	0.722	0.537	0.269	0.252	0.223	0.175
SD2	0.104	0.151	0.145	0.102	0.093	0.144	0.091	0.099	0.285	0.288	0.284	0.207	0.094	0.227	0.232	0.417	0.332	0.405	0.422	0.409	0.417	0.288	0.3	0.722	1	0.55	0.271	0.251	0.205	0.112
SD3	0.104	0.202	0.142	0.108	0.147	0.141	0.231	0.229	0.302	0.303	0.297	0.249	0.148	0.239	0.314	0.344	0.236	0.355	0.364	0.342	0.359	0.336	0.395	0.537	0.55	1	0.215	0.236	0.199	0.175
ITSM1	0.063	0.09	0.08	0.097	0.079	0.102	0.066	0.082	0.318	0.325	0.338	0.305	0.134	0.349	0.206	0.461	0.41	0.396	0.405	0.487	0.439	0.2	0.152	0.269	0.271	0.215	1	0.832	0.632	0.462
ITSM2	0.073	0.103	0.084	0.118	0.091	0.115	0.067	0.083	0.327	0.335	0.35	0.272	0.102	0.389	0.197	0.462	0.444	0.389	0.39	0.478	0.43	0.171	0.159	0.252	0.251	0.236	0.832	1	0.604	0.418
ITSM3	0.052	0.118	0.088	0.075	0.144	0.116	0.11	0.065	0.196	0.198	0.204	0.193	0.082	0.221	0.173	0.34	0.314	0.31	0.312	0.361	0.335	0.168	0.162	0.223	0.205	0.199	0.632	0.604	1	0.543
ITSM4	0.058	0.126	0.088	0.062	0.146	0.131	0.119	0.071	0.136	0.139	0.139	0.162	0.05	0.175	0.148	0.242	0.222	0.224	0.222	0.255	0.25	0.164	0.156	0.175	0.112	0.175	0.462	0.418	0.543	1

6.3.2 Correlations Coefficients

The correlation coefficient, a concept from statistics, is a measure of how well trends in the predicted values follow trends in past actual values. It is a measure of how well the predicted values from a forecast model "fit" with the real-life data. The correlation coefficient is a number between 0 and 1. If there is no relationship between the predicted values and the actual values the correlation coefficient is 0 or very low (the predicted values are no better than random numbers). As the strength of the relationship between the predicted values and actual values increases so does the correlation coefficient. A perfect fit gives a coefficient of 1.0. Thus the higher the correlation coefficient is the better. The equation for calculating the correlation coefficient is

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2} * \sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}} \qquad \text{(6-4)}$$

Where X_i and Y_i are the set of variables, $\overline{X}$ and $\overline{Y}$ the sample means.

We used the data shown in Table 6-2 and Table 6-3 to compute the correlation coefficient between the two tables. The result of correlation coefficient is equal to 0.670617. This value is good, so the relationship between the obtained results in chapter 5 and the obtained results of questionnaire is high. This indicates that the results of our method is near to reality. Moreover, the correlation coefficient between any special field in Table 6-2 and the same special field of the questionnaire, Table 6-3, was computed. Table 6-4 shows the result of field-field correlation coefficient.

Table 6-4: Field-Field Correlation Coefficient

Special Fields	Correlation Coefficient
Correlation between the document and the questionnaire	0.670617
R(M1,M1q)	0.837843
R(M2,M2q)	0.858454
R(M3,M3q)	0.88588
R(S1,S1q)	0.715951
R(S2,S2q)	0.943482
R(S3,S3q)	0.708698
R(C1,C1q)	0.752006
R(C2,C2q)	0.783699
R(ITA1,ITA1q)	0.821443
R(ITA2,ITA2q)	0.798621
R(ITA3,ITA3q)	0.830572
R(PM1,PM1q)	0.361256
R(PM2,PM2q)	0.550385
R(PM3,PM3q)	0.541665
R(PM4,PM4q)	0.477873
R(ITS1,ITS1q)	0.699458
R(ITS2,ITS2q)	0.84506
R(ITS3,ITS3q)	0.653267
R(ITS4,ITS4q)	0.606637
R(ITS5,ITS5q)	0.694207
R(ITS6,ITS6q)	0.588576
R(AS1,AS1q)	0.69571
R(AS2,AS2q)	0.754574
R(SD1,SD1q)	0.788018
R(SD2,SD2q)	0.773465
R(SD3,SD3q)	0.764054
R(ITSM1,ITSM1q)	0.796168
R(ITSM2,ITSM2q)	0.835193
R(ITSM3,ITSM3q)	0.89114
R(ITSM4,ITSM4q)	0.86881

The high correlation values means that the contents of the documents of ITSS is good and represent the reality. However, the low values mean that the documents of ITSS should be improved to reflect the reality. As example, correlation value for "Marketing" and "sales" is high, this indicates that the contents of them in ITSS is clear. However, project management has low values so, the content of them needed to be improved.

The obtained results can may help the Japanese ITSS team work to overcome and develop the weak points in the documents. The development can be done by cooperation between IT professionals, companies representatives, educational institute representatives and many others occupations that are interested in ITSS.

CHAPTER-7

Analysis of some Samples of ITSS's Career Paths

7.1 INTRODUCTION

This chapter analyzes some examples of desirable career paths for IT professionals. These examples of career paths were published in a document that was promoted by IPA. It is supposed that , changes in occupation are based on the assumption that the individual involved will remain at the same skill level before and after the change in occupation. In reality, an individual will generally descend to a lower skill level following a change in occupation. Moreover, it is assumed that the person must receive the necessary education and training in order to remain at the same level of skill.

7.2 PROPOSED METHOD

To analyze these career paths, I considered every level of every special field is as a text document. As we mentioned in the previous chapters, IPA has published ITSS as 3 parts. We formed these documents by extracting some sections from part 2 and part 3, then merging these sections for every level. For example to form the document of level 3 of the special field marketing:

- We extract the sections concerned to marketing's level 3 from the documents of ITSS's part 2 and part 3.
- We merge these sections to form only one document.

At this stage every level is represented by a text document. We used text mining techniques to analyze these documents. These techniques are as keyword extraction, document representation, term weighting, similarity computation.

1. For every document, extracting keywords using AnalogX keyword extraction tool.
2. Filtering keywords set by deleting the stop words.
3. Applying stemming by collapsing the words having the same root (e.g., activate, activates, activation, and active) to the same word for frequency counting.
4. Mathematically representing the documents using vector space model and weighting the keywords by the formula:

$$w_i = \frac{t_i}{\sum_{i=1}^{n} t_i}$$

Where, t_i is the frequency of i^{th} term in a document D.

5. Computing the cosine similarity between different documents.

$$sim(D_i, D_j) = \frac{D_i \times D_j}{\|D_i\| * \|D_j\|}$$

Where "$\times$" denotes the dot-product of the two vectors D_i and D_j and $\|D_i\|$ denotes the length or norm of a vector D_i .

We used the obtained values of cosine similarity to analyze the examples for career paths, promoted by IPA, that will be shown in the following section.

7.2.1 Career path: Marketing, Sales and Consulting

Figure 7-1 shows a career path for Marketing, Sales, and Consultant. By applying the proposed method, every arrow will have a value that represents the value of similarity measure as shown in Figure 7-2.

7.2.2 Career path: Software Product development

Figure 7-3 and Figure 7-4 show the career path for software product development before and after applying the proposed method.

7.2.3 Career path: Application development

Figure 7-5 and Figure 7-6 show the career path for application development before and after applying the proposed method.

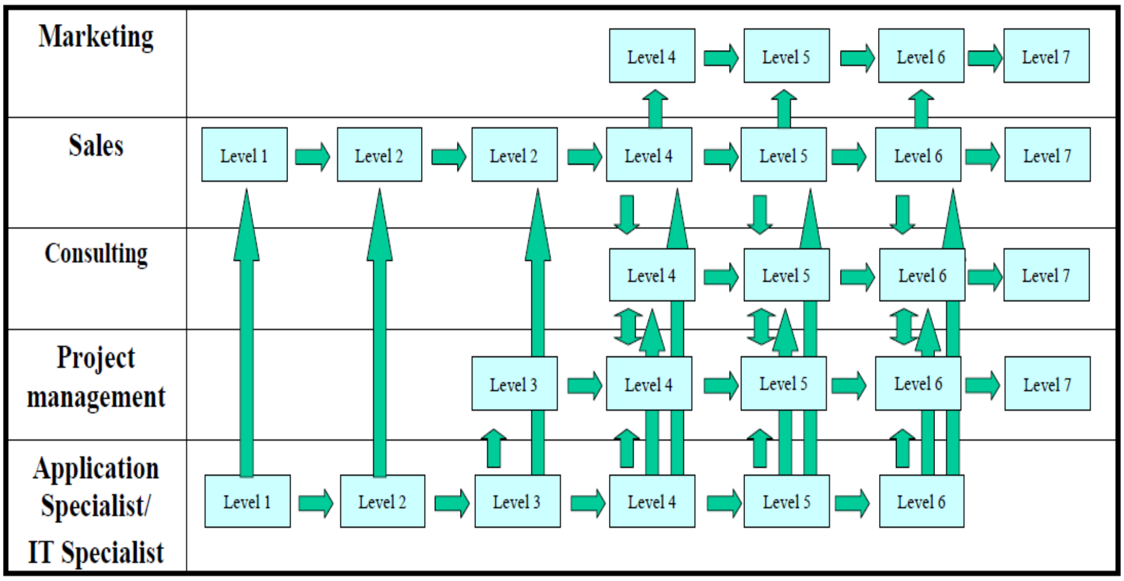

Figure 7-1: Marketing, Sales and Consulting career path

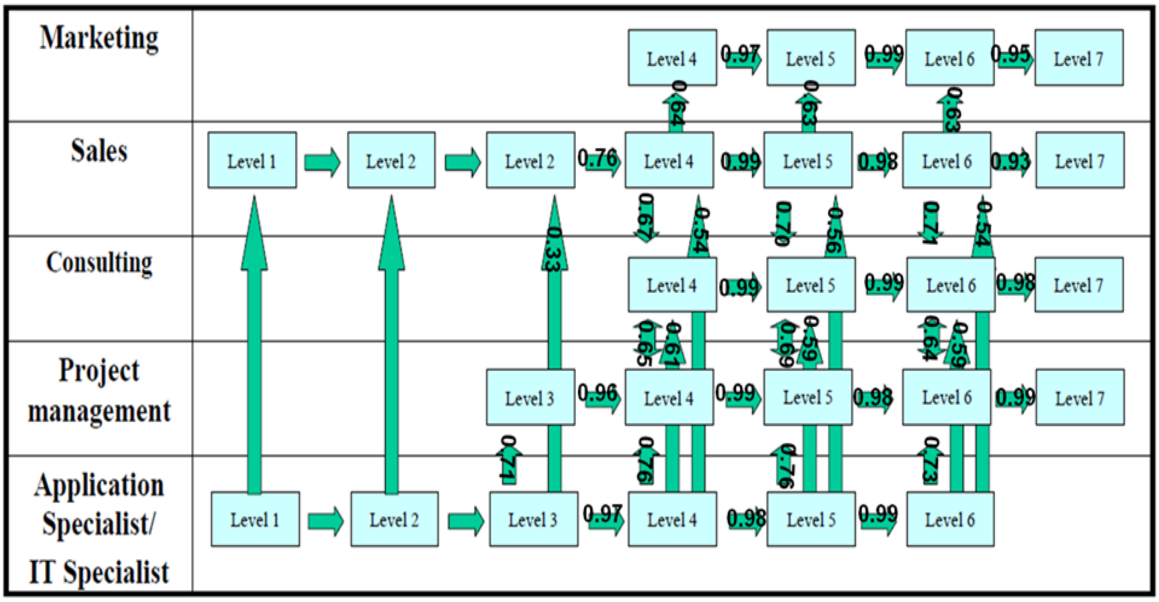

Figure 7-2: Career path after applying the proposed method

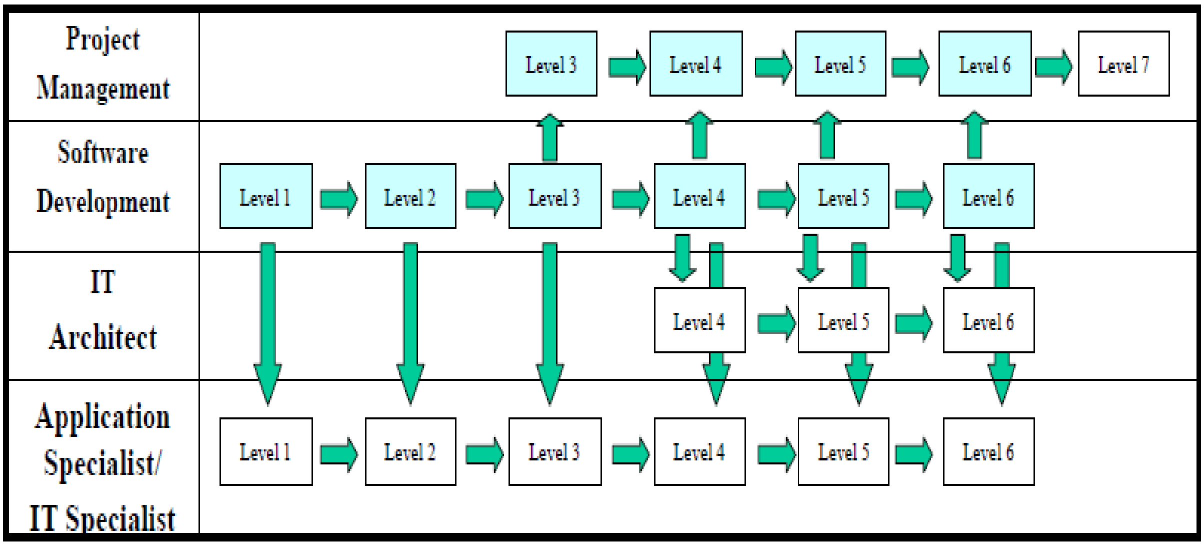

Figure 7-3: Software product development career path

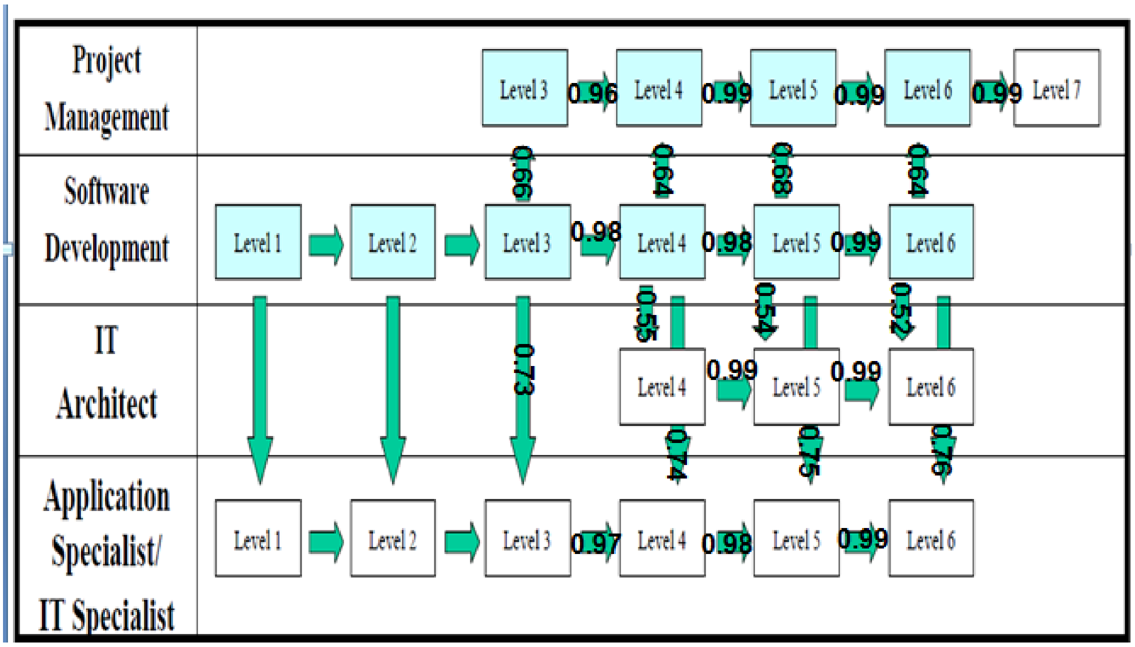

Figure 7-4: Arrows' values for the Software product development career path

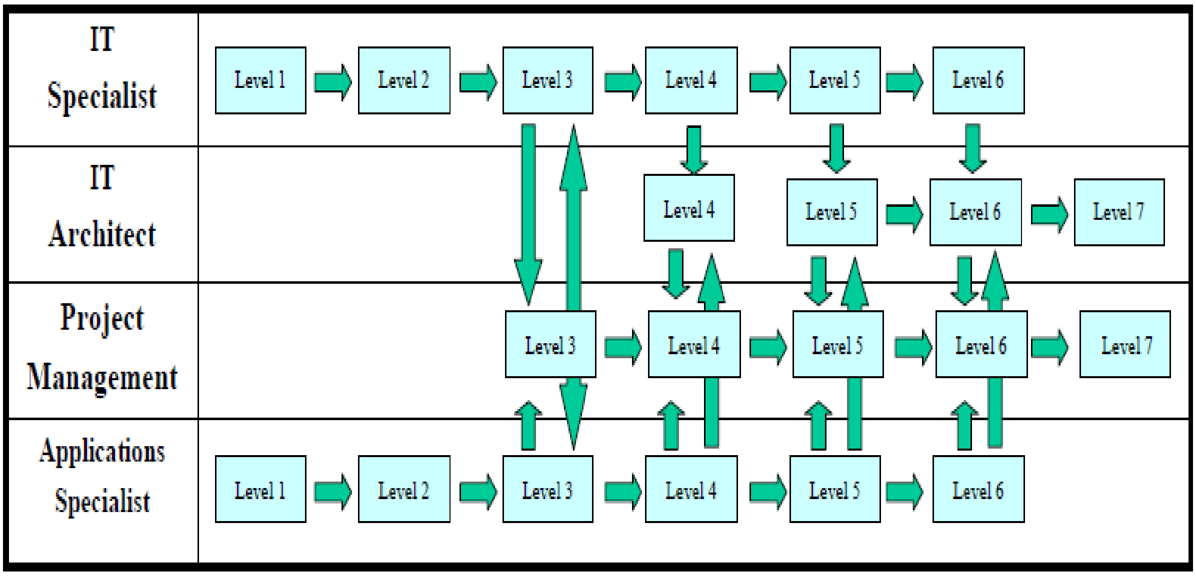

Figure 7-5: Application development career path

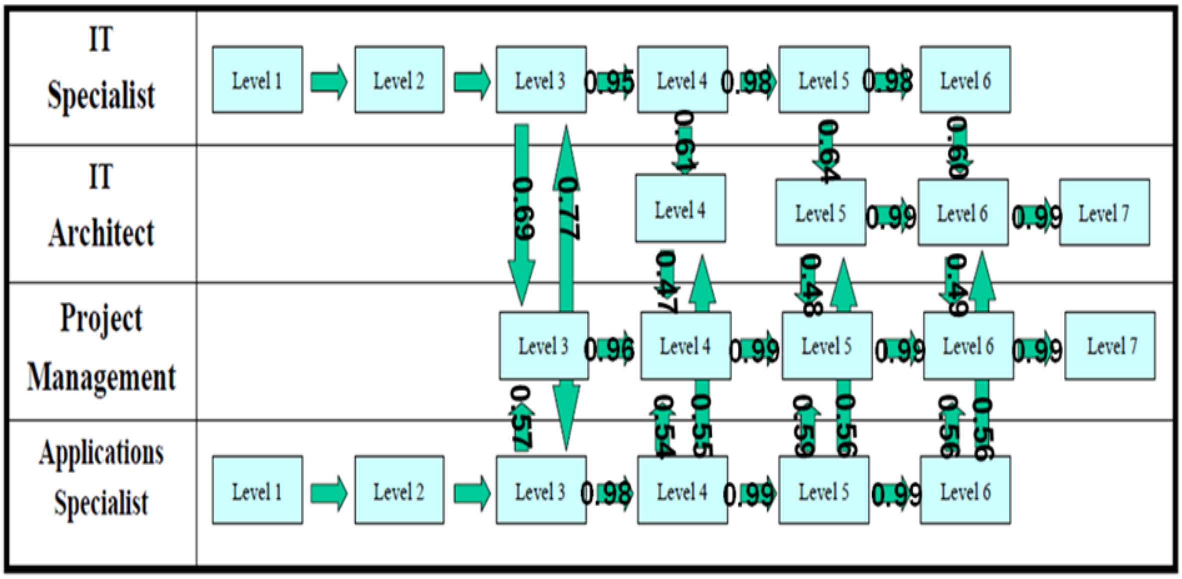

Figure 7-6: Arrows' values for the Application development career path

7.2.4 Analysis of the career path:

- To upgrade your level within the same job category requires less study than moving to another job category even you will move to the same level.
- If the IT engineer on level one or two, then the easiest way to be high level project manager is to upgrade himself to IT specialist level 3 then change to project management level 3 then skill up himself to the required level in project management. This is because Sim(IT specialist, Project Management). is greater than Sim(Application Specialist, Project Management). So, IT engineer will study less in this career path as shown in Figure 7-7 .
- If the IT engineer want to be a high level IT architect, then the easiest way is to go through IT specialist job category. Skill up to level 4 of IT specialist then move to IT architect category as shown in Figure 7-8.
- If an IT engineer wants to be a consultant, the easies path is to begin from sales job category till level 4 then transfer to consultant job category as shown in Figure 7-9 .

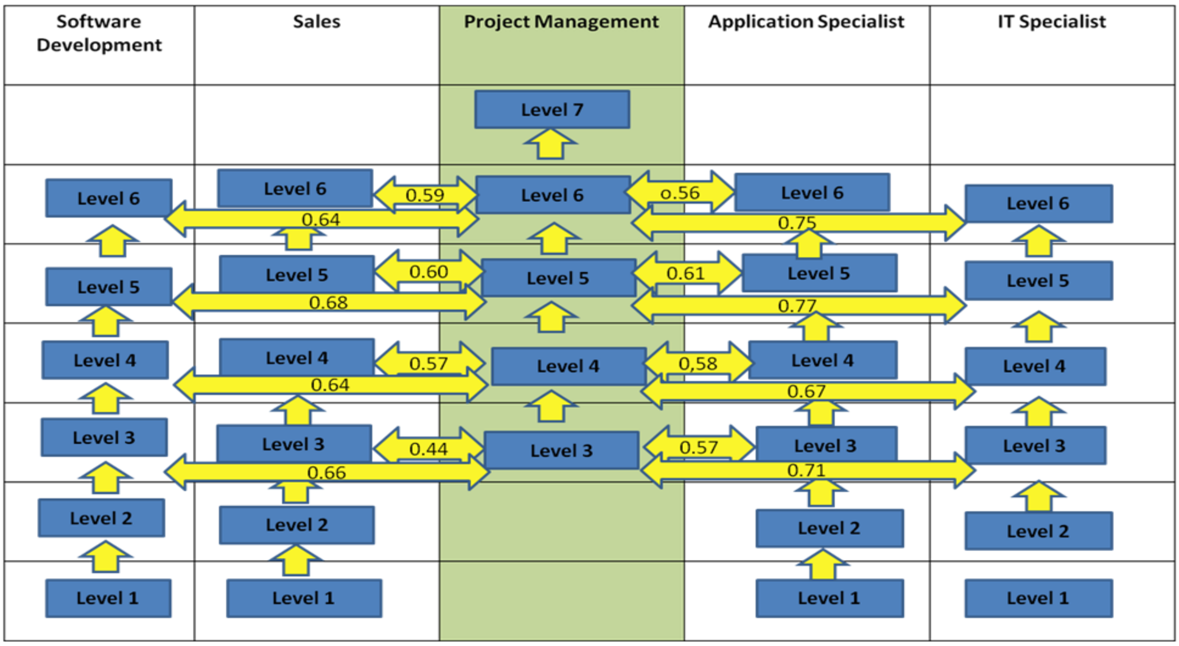

Figure 7-7: A career path for project management

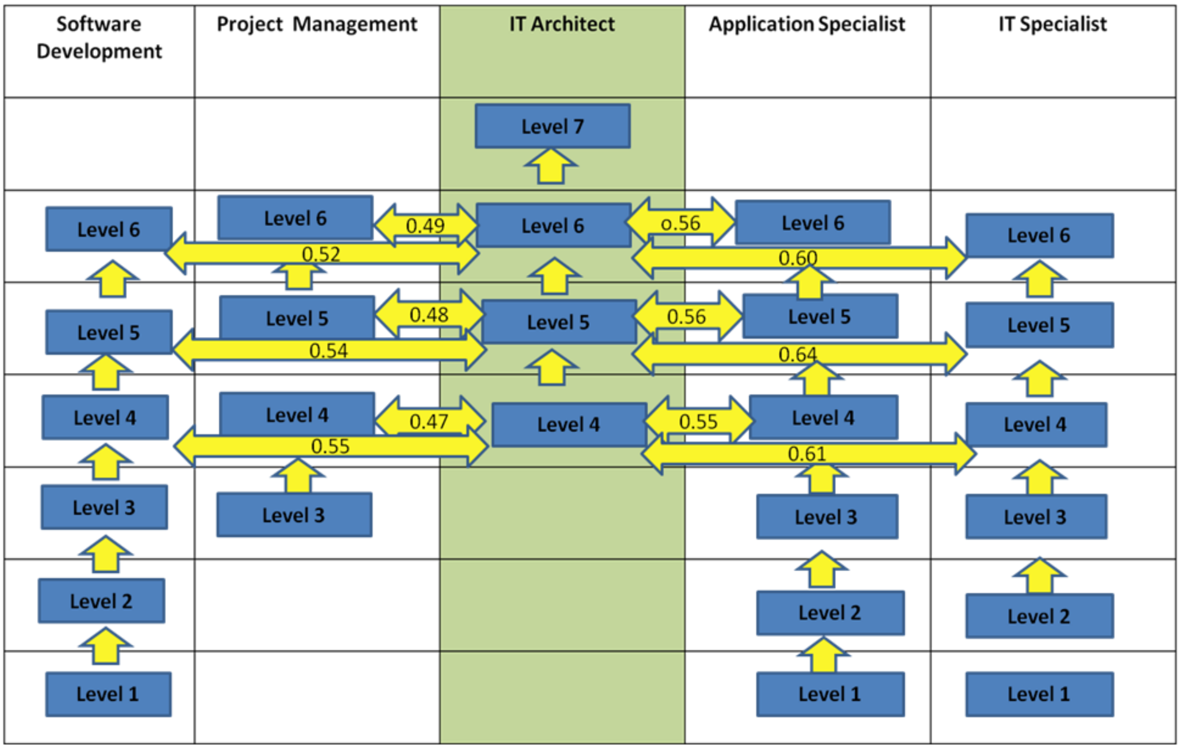

Figure 7-8: IT architect career path

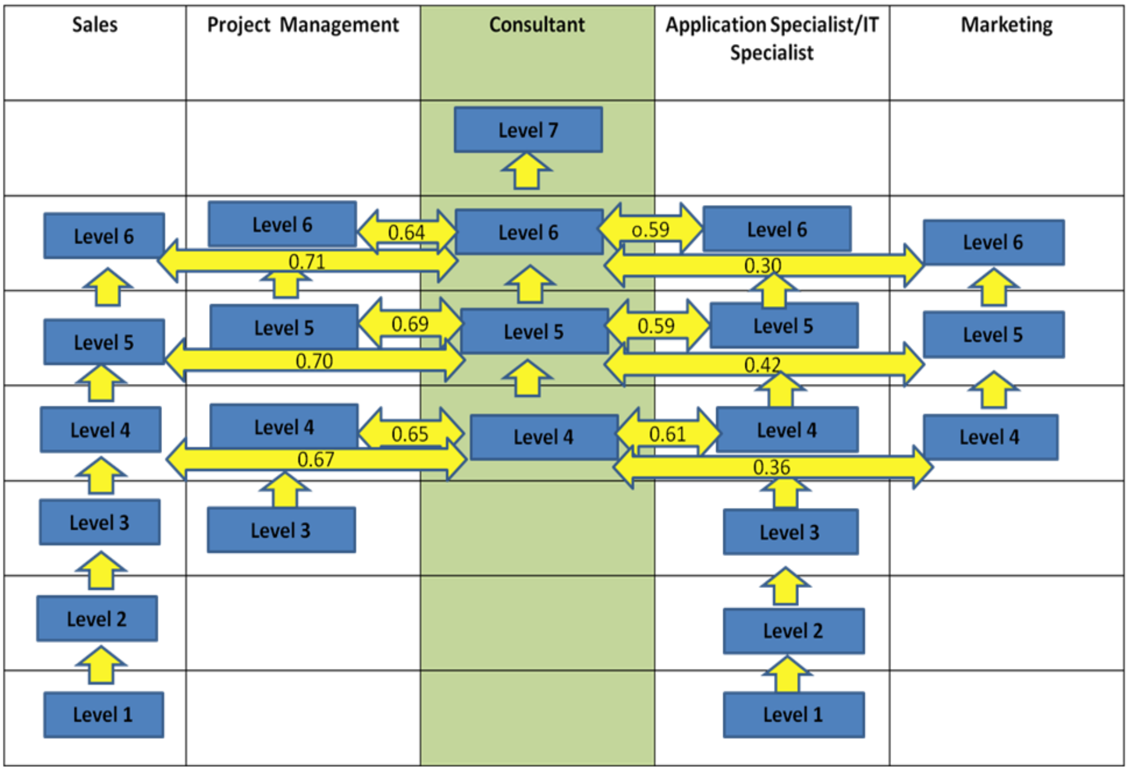

Figure 7-9: Consultant Career path

7.2.5 A proposed career path of IT architect:

From the earlier examples of career paths we proposed the following career path for IT architect as an example of typical model of a job category as shown in Figure 7-10.

In this career path an IT engineer can begin at the first level of application specialist, software development, or IT specialist. Then he can upgrade himself to level 3. From level 3 he can upgrade himself to level 4 then change to level 4 of IT architect, or, he can move to level 3 of project management then move to IT architect.

We supposed that IT engineer cannot change his career on level 6 and level 7. At level 6, it is better for him to move up to level 7 of the same job category. Moreover, at level 7 IT engineer becomes a world-class IT professionals. So, it is better to stay on his job category to lead advanced service development and commercialization for the whole IT market.

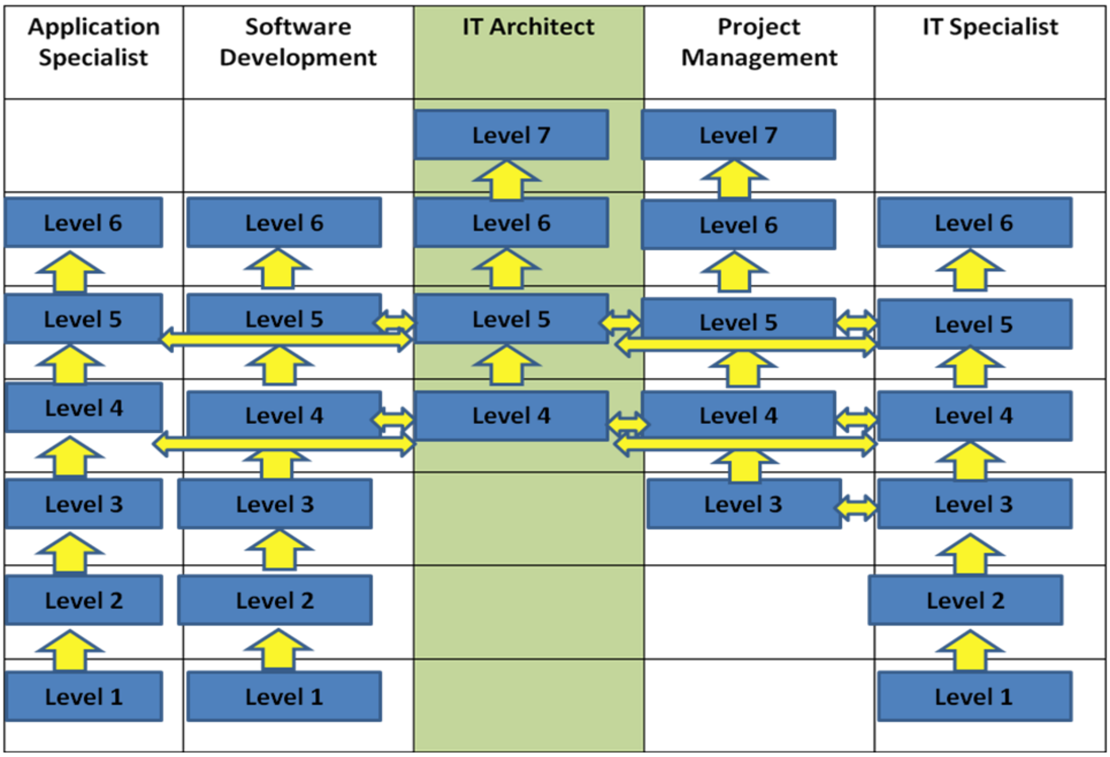

Figure 7-10: Proposed career path for IT architect

CHAPTER-8

Conclusion

With the rapid pace of developments in the business landscape, the industry has, more than ever, recognized the importance of effective human resource (HR). Developing a talented pool of IT professionals is crucial in sustaining business competitiveness. This is the premise behind the Information Technology Skill Standards (ITSS) developed by the Information Technology Promotion Agency (IPA) of Japan. This thesis analyzed some of the documents of ITSS that was published by IPA. The results of this study is important for both IT engineers and the developer of ITSS as some experts in software engineering said. Because according to the results of this study, the contents of the documents of some special fields and job categories should be improved to be more realistic. Moreover, the results can help IT engineers to decide what are the priority of the required courses or what are the required skills to upgrade themselves. This concluding chapter of the thesis is divided to two parts: summary of the findings and recommendation to further research as in the following sections.

8.1 SUMMARY FINDINGS

This thesis provided a basic methodology, questionnaire, text mining techniques, data analysis to analyze the documents of ITSS. The results will be helpful for IT engineers and for ITSS's developer as well. A summary of each chapter is discussed hereafter.

The first chapter was an introduction of this thesis.

The second chapter was the survey of this thesis. In this chapter the definition of information technology and skill standards has been discussed. In addition, we discussed the influence of the Internet and IT on education, IT market and the workforce. This chapter provided the previous research that showed the effect of IT and skill standards on human resources. Moreover, this chapter discussed who is should use the skill standards and explained the pyramid of skill standards. Furthermore, this chapter showed the Common Skills across all IT Human Resources. Finally, a brief description about some text mining techniques as document clustering, weighting schemes, keyword extraction, and similarity measures are presented.

The third chapter introduced a detailed description of the Japanese information technology skill standards. It started with explaining the reasons that the Japanese government published ITSS. Next, the human resource development process has been explained. This chapter included the description of the fundamental structure of ITSS with all the components. In addition the definition of some concepts as KPIs, skill proficiency, skill area, levels, complexity, size and responsibility, have been introduced. Also, the relationship between all the components of ITSS has been showed. Finally, the adoption process of ITSS, are explained which is very important for all the companies that will use ITSS.

The fourth chapter explained our methodology for the analysis of the eleven job categories of ITSS. Moreover, this chapter proposed a method to sort the required training courses list for each job category that were suggested by IPA.

The fifth chapter explained our methodology for the analysis of the 35 ITSS special fields. It provided a method extract the required keywords to move from one special field to another. In that method we formulated a weighting scheme for the common and special keywords for the special fields. In addition, the results were discussed by a group of software experts. The comments of the experts were presented in this chapter.

The sixth chapter explained a questionnaire that was applied to 1014 employees of a Japanese company to grasp the skill standards of the employees. The target of this chapter is to compare between the questionnaire's results and the obtained results in chapter 5. We changed the form of the data of the questionnaire to perform the comparison. Then the comparison was applied by computing the correlation coefficient between the questionnaire and the results of chapter 5.

The seventh chapter provided an analysis of some examples of career paths that were published by IPA. From the analysis process we suggested some points for IT engineer to be a high level project manager or consultant. Also, a proposed career path for IT architect was introduced .

The eighth chapter is the present chapter which concludes all the summary findings of this thesis.

8.2 RECOMMENDATIONS FOR FURTHER RESEARCH

This thesis focused on the Japanese ITSS by analyzing its documents . In future, we should compare in more details between ITSS and other standards as the American skill standards, the European framework, etc.. In addition we would like to study the probability of the combination between the not Asian standards and the Japanese standards ITSS. Also, we will analyze the adoption process of ITSS and study every step to extract some information that can help companies to undergo to this process easily.

Refrences

Adewoye J. O. and Obasan Kehinde A. The Impact of Information Technology (IT) on Human Resource Management (HRM): Empirical evidence from Nigeria Banking Sector. Case Study of Selected Banks from Lagos State and Oyo State in South-West Nigeria [Journal] // European Journal of Business and Management. - 2012. - 6 : Vol. 4. - pp. 28-38.

Allen Fred MANUFACTURING/ENGINEERING TECHNOLOGY SKILL STANDARDS. [Report]. - [s.l.] : The Boeing Company and the State of Washington through the State Board for Community and Technical Colleges, 1999.

AnalogX AnalogX Keyword Extractor [Online]. - AnalogX . - www.analogx.com.

ANTONIU ELIZA CAREER PLANNING PROCESS AND ITS ROLE IN HUMAN [Journal] // Annals of the University of Petroşani, Economics, 10(2),. - 2010. - 2 : Vol. 10. - pp. 13-22.

ANTONIU ELIZA CAREER PLANNING PROCESS AND ITS ROLE IN HUMAN [Journal] // Annals of the University of Petroşani, Economics. - 2010. - 2 : Vol. 10. - pp. 13-22.

Ardichvili A. Knowledge Management, Human Resource Development, and Internet Technology [Journal] // Advances in Developing Human Resources. - 2002. - 4 : Vol. 4. - pp. 451-463.

Ayad Hanan and Kamel Mohamed S. Topic Discovery from Text Using Aggregation of Different Clustering Methods [Conference] // AI '02 Proceedings of the 15th Conference of the Canadian Society for Computational Studies of Intelligence on Advances in Artificial Intelligence. - [s.l.] : Springer-Verlag London, UK, 2002. - pp. 161-175 .

Bailey S. and Morrell C. The implementation of Skills Framework for the Information Age (SFIA) at INACAP University Chile [Conference] // Embedding Skills Framework for the Information Age (SFIA) in the Curriculum, INACAP Technological University of Chile. - 2011.

Bal Yasemin The new human resources management in the 21st century: a strategic view [Conference] // Annual Conference on Innovations in Business & Management. - London, UK : [s.n.], 2011.

Baloh Peter and Trkman Peter Influence of Internet and Information Technology on Work and Human Resource Management [Conference] // Informing Science. - 2003. - pp. 498-505.

Bell Bradford S., Lee Sae-Won and Yeung Sarah K. The impact of eHR on professional competence in HRM : implications for the development of HR professionals [Article] // Working paper series / Center for Advanced Human Resource Studies. - June 2006.

Benson A.D., Johnson, S.D. and Kuchinke, K.P. The Use of Technology in the Digital Workplace: A Framework for Human Resource Development [Journal] // Advances in Developing Human Resources. - 2002. - 4 : Vol. 4. - pp. 392-404.

Bracewell D.B., Ren F. and Kuriowa S. Multilingual single document keyword extraction for information retrieval [Conference] // Natural Language Processing and Knowledge Engineering, 2005. IEEE NLP-KE '05. Proceedings of 2005 IEEE International Conference on. - 2005.

Breyer Jutta, Hook Terry and Marinoni Clementina Towards a European e-Competence Framework Interim Report", CEN/ISSS Workshop on ICT Skills CEN – European Committee for Standardization, 2007 [Report]. - 2007.

Brunzel Marko and Spiliopoulou Myra Domain Relevance on Term Weighting [Journal] // NLDB, Lecture Notes in Computer Science. - 2007. - Vol. 4592. - pp. 427-432.

Buckley Chris, Singhal Amit and Mitra Mandar New Retrieval Approaches Using SMART [Conference] // Proceedings of the Fourth Text Retrieval Conference (TREC-4). - 1995. - pp. 25-48.

Cajander Åsa [et al.] Assessing Professional Skills in Engineering Education [Conference] // Australasian Computing Education Conference (ACE 2011)CRPIT, 114.. - Perth,Austalia : John Hamer and Michael de Raadt Eds., ACS., 2011. - pp. 145-154.

Christopher D.M., Prabhakar R. and Hinrich S.R. An Introduction to Information Retrieval [Book]. - Cambridge, England : Cambridge University Press, 2009.

Cole-Avent and Gail A. An information technology perspective on the knowledge skills and support services that lead to technological fluency among college student affairs professionals [Book]. - ATHENS : UNIVERSITY OF GEORGIA, 2008. - Ph.D..

Conner J. and Ulrich D. Human resource roles: creating value, not rhetoric [Journal] // Human Resource Planning. - 1996. - 3 : Vol. 19. - pp. 38-49.

COOPE I.D. and PRICE C.J. GRID-BASED OPTIMIZATION METHODS [Conference] // ON THE CONVERGENCE OF GRID-BASED METHODS FOR UNCONSTRAINED OPTIMIZATION. - New Zealand. : [s.n.].

Damashek Marc Gauging similarity with n-grams: Language-independent categorization of text [Journal] // Science 267(5199). - 1995. - pp. 843-849.

Deerwester S. C. [et al.] Indexing by latent semantic analysis [Journal] // Journal of the American Society of Information Science. - 1990. - 6 : Vol. 41. - pp. 391-407.

Drucker P. The Essential Drucker [Book]. - Oxford,UK : Butterworh-Heinemann, 2001.

Dumais S. T. Latent semantic indexing (LSI) and TREC-2 [Conference] // In D. Harman (Ed.), The Second Text Retrieval Conference (TREC2). - [s.l.] : (National Institute of Standards and Technology Special Publication 500-215, 1994. - pp. 105-116.

eCF e-CF 2.0 CWA Part CWA Part III, Building the e-CF – a combination of sound methodology and expert contribution [Report]. - 2010.

EnglishStopWords RANKS.NL [Online] // English StopWords. - http://www.ranks.nl/resources/stopwords.html.

Evans Neil Chapter 2 Information Technology Jobs and Skill Standards [Book Section] // Technology Everywhere A Campus Agenda for Educating and Managing Workers in the Digital Age . - [s.l.] : Jossey-Bass, A Wiley Company, 2002.

Feng Yan, Zhang Xuesheng and Liu Liying rid-based methods for linearly equality constrained optimization problems [Journal] // Journal of Applied Mathematics and Computing. - 2007. - 1-2 : Vol. 23. - pp. 269-279.

Gardner Sharyn D Virtual HR: The impact of information technology on the human resource professional [Journal]. - 2003. - 2 : Vol. 63. - pp. 159–179.

Gardner Sharyn D., Lepak David P and Bartol Kathryn M. Virtual HR: The impact of information technology on the human resource professiona [Journal] // Vocational Behavior. - 2003. - 2 : Vol. 63. - pp. 159–179.

Haines III V.Y. and Lafleur, G. Information Technology Usage and Human Resource Roles and Effectiveness [Journal] // Human Resource Management. - 2008. - 3 : Vol. 47. - pp. 525-540.

Hamasuna Yukihiro Clustering Algorithms with various types of tolerance. - 2010.

Hawk Stephen [et al.] The Information Technology Workforce: A Comparison of Critical Skills of Clients and Service Providers [Journal] // Information Systems Management. - 2012. - pp. 2-12.

Hendrickson A. R. Human Resource Information Systems: Backbone Technology of Contemporary Human Resources [Journal] // Labor Research. - 2003. - 3 : Vol. 24. - pp. 381-394.

http://www.ipa.go.jp/english/humandev/forth_download.html Information Technology Promotion Agency [Online]. - http://www.ipa.go.jp/english/humandev/forth_download.html.

http://www.ipa.go.jp/index-e.html Information Technology Promotion Agency , Japan [Online] // IPA. - http://www.ipa.go.jp/index-e.html.

http://www.sfia-online.org/ [Online]. - http://www.sfia-online.org/.

Huang Anna Similarity Measures for Text Document Clustering [Conference] // proceedings of the New Zealand Computer Science Research Student Conference. - New Zealand : [s.n.], 2008. - pp. 49-56.

Huang Similarity Measures for Text Document Clusterin [Conference] // NZCSRSC. - Newsland : [s.n.], 2008.

Hulty A. Improved automatic keyword extraction given more linguistic knowledge [Conference] // EMNLP '03 Proceedings of the 2003 conference on Empirical methods in natural language processing. - [s.l.] : Association for Computational Linguistics Stroudsburg, PA, USA, 2003.

ILANGO and MOHAN V. A Survey of Grid Based Clustering Algorithms [Journal] // International Journal of Engineering Science and Technology. - 2010. - 2 : Vol. 8. - pp. 3441-3446.

Jain A. K. and Dubes R. C. Algorithms for Clustering Data [Book]. - [s.l.] : Prentice Hall, Englewood Cliffs, NJ., 1988.

Jeh Glen and Widom Jennifer SimRank: a measure of structural-context similarity [Conference] // KDD '02 Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. - [s.l.] : ACM New York, NY, USA , 2002. - pp. 538-543.

Jones K. Spark and Willett P. Chapter 3 [Book Section] // Readings in information

retrieval. - [s.l.] : Morgan Kaufmann Publishers Inc. San Francisco, CA, USA , 1997.
Jones Karen Sparck and Willett Peter Chapter 3 [Book Section] // Readings in information retrieval. - [s.l.] : Morgan Kaufmann Publishers Inc. San Francisco, CA, USA ©1997, 1997.
Jones Karen Sparck and Willett Peter Readings in information retrieval [Book]. - [s.l.] : Morgan Kaufmann Publishers Inc. San Francisco, CA, USA , 1997.
K.Sathiyakumari, V.Preamsudha and G.Manimekalai A Survey on Various Approaches in Document Clustering [Journal] // Int. J. Comp. Tech. Appl.. - 2011. - 5 : Vol. 2. - pp. 1534-1539.
Kondrak Grzegorz N-Gram Similarity and Distance [Conference] // Proceeding SPIRE'05 Proceedings of the 12th international conference on String Processing and Information Retrieval. - [s.l.] : Springer-Verlag Berlin, Heidelberg, 2005. - pp. 115-126.
Krishna Murali and Bhavani S.Durga An Efficient Approach for Text Clustering Based on Frequent Itemsets [Journal] // European Journal of Scientific Research. - 2010. - 3 : Vol. 42. - pp. 399-410.
Lan M. [et al.] (2005) A comprehensive comparative study on term weighting schemes for text categorization with support vector machines. [Conference] // Posters proceedings of 14th international world wide web conference. - [s.l.] : ACM New York, NY, USA, 2005. - pp. 1032–1033.
Landauer Thomas K. [et al.] Handbook of Latent Semantic Analysis [Book]. - [s.l.] : University of Colorado Institute of Cognitive Science Series, 2007.
Lin Liang-Hung Electronic human resource management and organizational innovation : the roles of information technology and virtual organizational structure [Journal] // The international journal of human resource management. - London, Routledge : [s.n.], 2011. - 2 : Vol. 22. - pp. 235-257.
Matsuo Y. and Ishizuka M. KEYWORD EXTRACTION FROM A SINGLE DOCUMENT USING WORD CO-OCCURRENCE STATISTICAL INFORMATION [Journal] // International Journal on Artificial Intelligence Tools. - [s.l.] : International Journal on Artificial Intelligence Tools, 2004. - 1 : Vol. 13. - pp. 157-169.
McClure and Ann Polley Chapter 1: Managing the Complexity of Campus Information Resources [Book Section] // Organizing and Managing Information Resources on your Campus / book auth. McClure Polley Ann. - [s.l.] : Jossey-Bass, Willey company, 2003.
Mishra Alok and Akman Ibrahim InformationTechnology in Human Resource Management : An Empirical Assessment [Journal] // Public Personnel Management Volume. - 2010. - 3 : Vol. 39. - pp. 243-262.
Misra Harekrishna Role of Human Resource in Information Technology Alignment in Organizations: A metric Based Strategic Assessment Framework [Journal] // Information Technology Management. - 2006. - 3 : Vol. XVII. - pp. 38-50.
Murugesan Keerthiram and Zhang Jun A NEW TERM WEIGHTING SCHEME FOR DOCUMENT CLUSTERING [Conference] // IEEE International Conference on Data Mining

- ICDM. - 2011.
Nazari Kamran [et al.] The Role of Information Technology in Human Resource [Journal] // Basic and Applied. - 2012. - 2 : Vol. 6. - pp. 5447-5451.
NCR National Research Council, Computer Science and Telecommunications Board (2001) Building a Workforce for the Information Economy [Report]. - Washington, DC : National Academy Press, 1999.
Newsletter The European e-Skills Newsletter: Issue No 4 [Report]. - 2005.
NWCET Building a foundation for tomorrow: Skill standards for information technology [Report]. - Bellevue : WA: Bellevue Community College, 2003.
NWCET New designs for advanced information technology education. [Report]. - Bellevue : (Renewal Grant to NSF). WA: Bellevue Community College., 1997.
NWCET Building a foundation for tomorrow [Report]. - Bellevue : Bellevue Community College Printing Services. , 1999.
NWCET, RATEC and BCC Building a foundation for tomorrow: Skill standards for information technology. (Progress Report). [Report]. - Bellevue : NWCET, RATEC & BCC, 1996.
Porter M. F. An algorithm for suffix stripping [Book Section] // Readings in information retrieval. - [s.l.] : Morgan Kaufmann Publishers Inc. San Francisco, CA, USA ©1997, 1997.
Recruiting Global 500 Web Site iLogos Research, San Francisco. [Report]. - San Francisco : [s.n.], 2000.
Reed J. [et al.] TF-ICF A New Term Weighting Scheme for Clustering Dynamic Data Streams [Conference] // 5th International conference on Machine Learning and Applications. - 2006. - pp. 258-263.
Reynolds A. P. [et al.] Clustering Rules: A Comparison of Partitioning and Hierarchical Clustering Algorithms [Journal] // Journal of Mathematical Modelling and Algorithms. - 2006. - 4 : Vol. 5. - pp. 475-504.
RIS Raba Interneta v Sloveniji . Rezultati raziskave o rabi Interneta v podjetjih, [Report]. - 1999. - Retrieved from http://www.ris.org.
Rokach Lior and Maimon Oded 14 A survey of Clustering Algorithms [Book Section] // DATA MINING AND KNOWLEDGE DISCOVERY HANDBOOK. - [s.l.] : Springer New York Dordrecht Heidelberg London, 2010.
Salton G. and Buckley C. Term-weighting approaches in automatic text retrieval [Journal] // Information Processing and Management. - 1988. - 5 : Vol. 24. - pp. 513-523.
Salton G. Automatic text processing: the transformation, analysis, and retrieval of information by computer [Book]. - [s.l.] : Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA, 1989.
Salton G. The SMART Retrieval System—Experiments in Automatic Document Processing [Book]. - [s.l.] : Prentice-Hall, Inc. Upper Saddle River, NJ, USA , 1971.
Salton G., Wong A. and Yang C. S. A vector space model for automatic indexing [Journal] //

Communications of the ACM. - 1975. - pp. 613-620 .
Salton Gerard Automatic Text Processing. [Book]. - New York : Addison-Wesley, 1989.
Sharma Rajeshr Kuma and Sharma D. P. Impact of Information Technology to Enhance Human Skills [Journal] // Current Trends in Technology and Science. - 2012. - 2 : Vol. 1. - pp. 71-73.
SIGITE Computing Curricula Information Technology Volume [Report]. - [s.l.] : IT Curriculum Committe, 2005.
Silva Sriyan de HUMAN RESOURCES DEVELOPMENT FOR COMPETITIVENESS: A PRIORITY FOR EMPLOYERS [Conference] // ILO Workshop on Employers' Organizations in in Asia-Pacific in the Twenty-First Century. - Turin, Italy : [s.n.], 1997.
Stracke Christian M. Competences and Skills in the Digital Age: Competence Development, Modelling, and Standards for Human Resources Development [Article] // CCIS 240. - 2011. - pp. 34-46.
Strohmeier Stefan Research in e-HRM: Review and implications [Journal] // Human Resource Management Review. - 2007. - 1 : Vol. 17. - pp. 19–37.
Survey Achieving Results with Internet Recruiting , iLogos Research [Report]. - San Francisco : [s.n.], 1998.
Taghva Kazem and Veni Rushikesh Effects of Similarity Metrics on Document Clustering [Conference] // ITNG '10 Proceedings of the 2010 Seventh International Conference on Information Technology: New Generations. - [s.l.] : IEEE Computer Society Washington, DC, USA, 2010. - pp. 222-226 .
Ulrich D., Younger, J. and Brockbank, W. The Twenty-First Century HR Organization [Journal] // Human Resource Management. - 2008. - 4 : Vol. 47. - pp. 829-850.
ULRICH DAVE A new mandate for human resources. [Journal] // HARVARD BUSINESS REVIEW. - 1998. - pp. 124-34.
Xi W. [et al.] SimFusion: measuring similarity using unified relationship matrix [Journal] // SIGIR. - 2005. - pp. 130-137.
Xu Rui and Wunsch D. II Survey of clustering algorithms [Journal] // IEEE Transactions on Neural Networks. - [s.l.] : IEEE Press Piscataway, NJ, USA, 2005. - 3 : Vol. 16. - pp. 645-678.
Yin Xiaoxin, Han Jiawei and Yu Philip S. LinkClus: Efficient Clustering via Heterogeneous Semantic Links [Journal] // VLDB '06. - [s.l.] : ACM., 2006.
ZHANG Chengzhi [et al.] Automatic Keyword Extraction from Documents Using Conditional Random Fields [Journal] // Journal of Computational Information System. - [s.l.] : Binary Information Press , 2008. - pp. 1169-1180.
Zhang Chengzhi [et al.] Document Clustering Description Extraction and Its Application [Conference] // ICCPOL '09 Proceedings of the 22nd International Conference on Computer Processing of Oriental Languages. Language Technology for the Knowledge-based Economy. - Berlin, Heidelberg : Springer-Verlag , 2009. - pp. 370 - 377 .
Zhong Shi and Ghosh Joydeep A Unified Framework for Model-based Clustering

[Journal] // Journal of Machine Learning Research 4 (2003). - 2003. - pp. 1001-1037.