Asia Pacific Journal of Information Systems
Vol. 29 No. 4 (December 2019), 717-730

ISSN 2288-5404 (Print) / ISSN 2288-6818 (Online)
https://doi.org/10.14329/apjis.2019.29.4.717

IT Jobs in the Era of Digital Transformation: Big Data Analytics

Ho Lee[a], Jaewon Choi[b,*]

[a] Assistant Professor, Department of Future Technology, Korea University of Technology and Education, Korea
[b] Assistant Professor, Department of Business Administration, Global Business School, Soonchunhyang University, Korea

A B S T R A C T

The era of digital transformation (or the fourth industrial revolution) has been triggered by the rapid development of software (SW) technologies. In this era, several studies suspected rapid changes in job structures occurring around the world. Thus, there is a growing need for acquiring the skill sets required for the future. However, there are no specific studies on how existing jobs are changing. To cope with this ambiguity of job changes, this paper aims to investigate how the current job structure is changing in response to digital transformation. To identify the dynamic nature of job change over time, we conducted an analysis based on job posting data. As a result, nine job occupations and fifteen jobs were found.

Keywords: Digital Transformation, New Jobs, Job Change, Job Classification, Big Data, Lda

Ⅰ. Introduction

The era of digital transformation (or the fourth industrial revolution) has been triggered by the rapid development of software (SW) technologies such as artificial intelligence, big data, internet of things (IoT), etc. In this era, not only the economy and society, but also everyday life and individual work styles are expected to undergo dramatic changes. In particular, changes in the job sectors are expected to be greater than ever. The World Economic Forum (2016) argues that "65% of children entering primary school today will ultimately end up working in completely new job types that do not yet exist."

This implies not only the creation of new jobs, but also changes in existing ones. To cope with the rapid changes in job structures occurring around the world, it is imperative to identify changes in domestic jobs. Frey and Osborne (2013) argued that 47% of American jobs are jobs that could be replaced or altered by the development of computers. Surprisingly, some studies argued that in Korea 63%

*Corresponding Author. E-mail: jaewonchoi@sch.ac.kr Tel: 82415301240

of jobs were affected by the development of computers (Lee, 2015), as the aftermath of digital transformation can be even bigger for Korea, which has more simple office jobs than the United States, the United Kingdom, and China. The World Economic Forum (2016) predicted that, by 2020, about 7.14 million jobs would be lost and about 2 million would be created. During rapid job changes, some people worry about a job crisis, while others consider it an opportunity, as new jobs created by technology (the emergence of new jobs and industries) can offer new opportunities (Frey et al., 2016). Rapid changes in jobs due to digital transformation are not concerns for the future, but what is already happening now. Thus, there is a growing need for acquiring the skill sets relevant in the future (Bakhshi et al., 2017); however, there are no specific studies on how existing jobs are changing.

Hence, this paper aims to find out how current job structures are changing in response to the digital transformation. Information Technology (IT) jobs may show the most significant changes, since digital transformation was triggered by IT technologies in the first place. Consequently, this paper focuses on changes in IT jobs. To identify job changes, job postings were classified using clustering techniques, while job characteristics were identified based on the words, which can be considered as the tasks that make up the cluster. Based on these characteristics, the names and definitions of jobs were defined according to the opinions of experts.

Ⅱ. Literature Review

2.1. Limitations in Current Job Classification Systems

Related concepts and classifications were analyzed to predict job transformation. A job occupation can be divided into job type, job, and tasks. A job type is a bundle of jobs with similar characteristics (Igbaria et al., 1991). A job, which is directly related to this study, is assembled and organized by various tasks performed in the profession, with tasks defined as various activities performed in a single job such as writing a report, presentation, and so on.

There are also various criteria for classifying jobs. First, there is the standard industrial classification based on industrial activity by production units that is generally used in various fields such as national economic analysis and outlook (Gardner, 2005). It is a classification based on units of production; thus, it is deemed inappropriate to analyze jobs. Second, there is the standard occupational classification that is directly related to this study, divided into 'the types of work that individuals are doing'. In general, jobs have changed slowly, but continuously, even if not due to the digital transformation (Dudley et al., 2006). Hence, the standard occupational classification has been revised continuously. Recently, the domestic (Korean) standard occupational classification (KSCO) was revised at July 17, 2017 by the National Statistical Office in Korea. However, such revisions were made based on the traditional classification system, which does not consider convergent jobs from multi-industries. In addition, the traditional standard occupational classification cannot identify emerging jobs, which do not have enough size to be distinguished, despite generating high demands from the industry. Therefore, the criteria for each job and job occupation should be updated according to the characteristics and changes in the industry (Apte et al., 2008). Additionally, this paper aims to analyze the actual demands by industry and derive occupational classifications and characteristics accordingly.

To overcome such limitation, we adopt a big data analysis method. Since job postings show the current situation in an industry, classifying emerging jobs through job posting analysis can reveal the characteristics of job changes and cope with the drastic changes in the era of digital transformation.

2.2. Previous Studies on Job Changes

Previous studies have investigated job changes using quantitative and qualitative analyses, with most quantitative studies focusing on changes in the number of jobs due to changes in the external environment (Bakhshi et al., 2017; Vogler et al., 2014). For instance, the World Economic Forum (2016) have estimated that seven million jobs would be lost worldwide over five years, with only about two million new jobs being created.

Similarly, qualitative studies have focused on changes in the listing of jobs. For example, a research by McKinsey Global Institute (2017) analyzed the effects of new technologies such as automation, robotics, and artificial intelligence on employment and productivity by considering 2,000 activities of about 800 jobs in all industries.

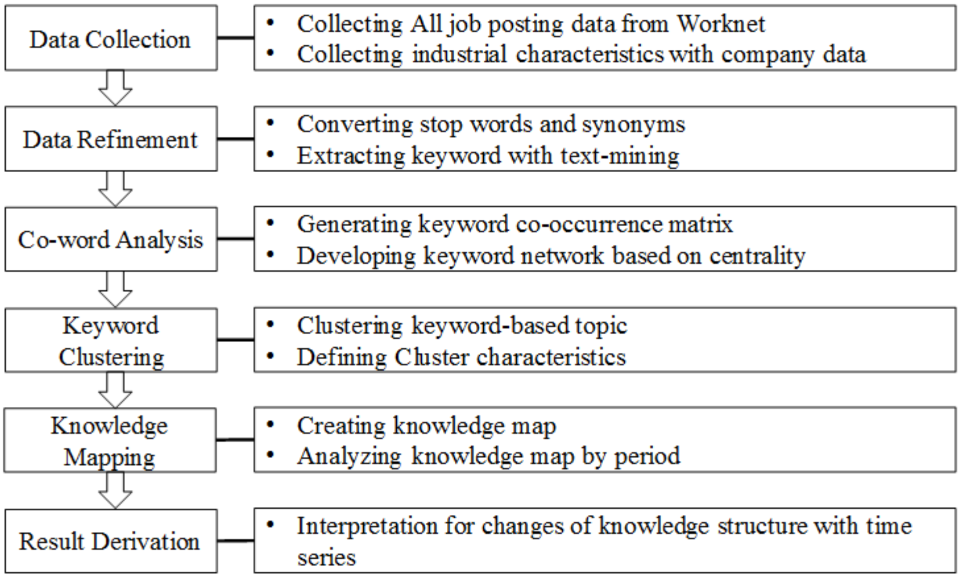
However, these studies are based on surveys or insights from experts, and did not analyze industrial data, considering jobs as static objects. These limitations can also be found in future job studies. Jobs change over time, as they have dynamic characteristics. In the same job, your tasks can be completely different as time goes by. For example, in the 80s, accountants did not need to know how to use a computer; however, after 10 years, they can be replaced by artificial intelligence. To detect the rapid change in an industry and to consider the dynamic nature of job change over time, we conducted an analysis based on job posting data.

Ⅲ. Research Methodology

The purpose of this study is to develop a job classification system based on data analysis derived from recruitment information. Accordingly, Worknet employment data were used to analyze recruitment announcements using text-mining techniques and obtain data by word to identify the job characteristics.

The data were collected and utilized from employment data published on Worknet, which is a website that provides job information, job search, and job search information, and is officially operated by the Ministry of Employment and Labor and the Korea Employment Information Service. It intuitively provides users with the information they need to develop a Korean work dictionary in accordance with the Korean occupational classification standards. Worknet materials are highly reliable and are judged to be a popular channel for both businesses and users. Therefore, searching the Worknet data and analyzing the data in the SW field are useful for identifying the jobs and job competency systems where the employment demand and workforce are concentrated, which are highly required. Data were collected from September 1, 2016 to December 25, 2016, totaling 20,408 information items. The data on recruitment announcements were collected and pretreated by word using text-mining techniques. Topic modeling based on the latent Dirichlet allocation (LDA) algorithm was conducted to derive a new classification system for tasks and professions (Doan et al., 2011; O'Reilly, 2005).

This study adopted a research procedure, as shown in <Figure 1>, for analysis through text mining. Effective words were derived through database construction and data clean-up. Afterwards, we conducted the correlation analysis between the relevant words, forming a cluster of derived words.



<Figure 1> The Process of Data Analysis

Consequently, the study was conducted in the order in which the words were identified, finding the relevant results (Pang and Lee, 2008).

In the data pre-processing stage, a web crawler was developed to convert unstructured data into structured data. The text within the recruitment notice was analyzed in natural language using the Python language. This method parsed words, filtered them, and quantified duplicate nouns. Afterwards, the top 50 words were compiled into data in the most common order. Subsequently, all possible errors in the data were excluded by developing a dictionary of terms related to 'computers and SW' among additional parsed nouns. In the data collection, ICT-related dictionary-based SW vocabulary dictionaries were created, so that pre-processed data could be obtained and topic modeling could be performed.

The data were collected using Python and scrapy/.xpath. We derived the matrix data based on the notion that the recruitment notice document consists of a set of related words. The process of deriving the frequency of words contained in a document within each cell through a matrix of words-documents (term-document matrices) consisting of the interrelationships between the documents in each condition is as follows. For the purpose of automating the classification of documents, a 'word frequency weight' was used to calculate the frequency of words, using words representative of the contents of each document and the method to determine the weight of those words. The term frequency weighting is defined as the proportion of a particular term to the entire document. By combining the frequency weights and term weights extracted from documents and words, relevant information can be obtained within and between the employment announcement documents.

After collecting and cleaning up the data, LDA was used to derive the keywords in the recruitment announcement to better predict the demand for the SW industry. LDA is an algorithm used in text-mining techniques that searches for topics hidden in docu-

<Figure 2> Cralwing Data and NLP

ments (Blei et al., 2003; Maybury, 2004). Compared to the theoretical classification methods identified by the existing classification systems, the LDA can identify the importance of the keywords (capabilities in the task) required by the industry for each task, allowing us to see the degree of change in SW workforce demand over time (Chen et al., 2012). The basic assumptions of the LDA model applied in this study are as follows: (1) There are k-topics and M documents in the total recruitment data collected, with all employment disclosure documents having a topic for a single employment; (2) Each employment announcement document shall appear in succession to N words and may be expressed as w_{dn}.

To derive the topic, the following steps were taken: (1) Poisson distribution: we used a probability distribution that indicates the frequency of occurrence of a particular event within a unit time to select an arbitrary N; (2) We selected Θ according to the Dirichlet distribution, which indicates of which topic(s) each employment announcement document consists; (3) For individual words for N words, (3_1) we selected topics according to the polynomial distribution of Θ; (3_2) we obtained words from when the topics are given. In other words, a specific topic, Θ, that follows the Dirichlet distribution of the entire employment announcement document, is derived and modelled by selecting the words classified in the document from the vector of the topic.

A word dictionary has been developed in a recruitment notice for the analysis of this study, with the vocabulary used containing 1,419 words. The process of developing a vocabulary dictionary was as follows. Words that appeared in the dictionary of terms in the software field were used for the composition of vocabulary. English vocabulary used words in TechTerms' IT-related dictionaries that were compiled in the categories 'Internet', 'HW', 'SW', and 'Technology'. Although the SW-related words mainly use abbreviations in English, Korean vocabulary was included in the vocabulary dictionary as the official usage term through the software term dictionary of the Korean Copyright Commission.

Five experts in the SW industry (three engineers and two academics) independently selected all words related to SW jobs and tasks, among the total 8,000 words extracted from the Worknet SW sector recruitment announcements, more than 50 words appeared. To ensure the reliability of the analysis results, the five coders were trained for about one hour to understand the research and match the criteria for word classification. This stage is a proactive measure that must be performed prior to the analysis to ensure consistency in the results of the qualitative analysis between the coders by conducting an intercoder reliability survey. We used the P/L index of Perreault and Lawrence (1989) to measure intercoder reliability. According to the intercoder relativity analysis, we found that the composition of the lexicon issued was significant, with 79.4% of the reliability of the vocabulary dictionary developed with the word list selected by the five experts. Therefore, 1,419 words were chosen as vocabulary dictionaries for analysis.

Ⅳ. Results

In this study, vocabulary dictionaries were organized and Worknet data analyzed based on words extracted from Worknet. Based on the word dictionary, the probability values in the matrix were calculated for the word co-occurrence of all words in the vocabulary dictionary. Word co-occurrence refers to the ratio of the frequency at which certain words in a topic appear simultaneously with other words in a recruitment notice document. The greater the word co-occurrence of a particular word, the more likely it is to be the 'keyword' that appears within that topic.

Topic modeling with each cluster feature was performed, as the words and announcements close to the distance of the overall probability distribution formed a cluster. Keywords for a total of fifteen jobs, nine job occupations, and each job or job occupations were derived through the LDA. A description of the LDA-based classification process is as follows. We conducted a focus group interview and decided upon an optimal number of fifteen clusters to determine the optimal number of job occupations. To verify the reliability of the interview results, we measured significance of the cluster results through intercoder reliability.

Since a decision on the number of clusters should be made first, the optimal cluster was determined for the top ten keywords of the group, which were classified from 5 to 30, based on the word co-occurrence extracted from Worknet. To this end, five experts on the SW industry (article: three; academic: two) were interviewed independently on their respective results. To verify the reliability of the interview results,

<Table 1> Research Contents for Current and Future Jobs

	Quantitative Analysis	Qualitative Analysis
Current Job	2017 Korea Occupational Outlook (Korea Employment Information Service, 2017) The future of jobs: Employment, skills and workforce strategy for the fourth industrial revolution (World Economic Forum, 2016) The German Labour Market in the Year 2030 (Vogler, Kurt, et al, 2014) The future of skills: Employment in 2030(Bakhshi, Hasan, et al., 2017) The future of employment (Frey and Osborne, 2013)	The risk of automation for jobs in OECD countries (Arntz, Melanie, Terry, et al., 2016) Artificial Intelligence, Automation, and the Economy (Executive Office of the President, 2016) A future that works: AI automation employment and productivity (McKinsey Global Institute Research, Tech. Rep, 2017).
Future Job	Predicting labor demand in future industry(Korea Institute for Industrial Economics & Trade and Korea Institute for Advancement of Technology, 2016) The gold vein of future jobs, Software (Won-Young Cho, Dong-hyun Lee, 2016)	New job research through domestic and overseas job comparison analysis (Korea Employment Information Service, 2016)

an intercoder reliability survey was conducted similar to that of a vocabulary dictionary, showing that the intercoder reliability for the cluster was 0.853 average of the five, resulting in 85.3% concordance.

LDA-based keywords (based on the co-existence frequency ratio) were extracted from the optimal clustering and higher tier clusters (fifteen jobs, nine job occupations.). In the case of job occupation, the same LDA was applied, analyzed, and classified into nine jobs as shown in <Table 1>. Based on the co-occurrence frequency of specific words in a cluster, the most important key words out of the total were derived and the cluster-specific characteristics were identified accordingly.

Moreover, words with a simultaneous frequency of less than 2% of each word were excluded from further analysis, since their influence on the entire cluster was minimal. In addition, a co-appearance word network was created based on social network analysis to find out the relationship between the words in the cluster. The co-appearance job network is a joint appearance of words in frequency, generated simultaneously with the relationship between words.

To obtain the task name and definition of each cluster extracted from <Table 2>, it was used to develop a job-and-job occupation system by referring to three international standards for job classification. The international classification of job standards system referred to in order to derive the classification system considers the following: (1) the international standard classification of occupation (ISCO)[1] by the International Labour Organization (ILO), (2) the standard occupational classification 2010 (SOC)[2] by the United States Bureau of Laboratory Statistics, and (3) the occupational information net (O*Net)[3] by the Employment and Training Administration of the United States Department of Labor. SOC-2016 was considered first, especially if there was a classification suitable for SOC-2016 as a preliminary revision of the most recent version of SOC.

First, to identify the job titles and definitions developed by LDA results, five experts in the SW field (two professors in SW-related departments and three industrial experts in SW-related fields) participated in the intercoder reliability analysis to ensure the suitability of the results. Each of the five selected experts (coder) examined the results by comparing the results with the top ten keywords of the word co-occurrence in the cluster separated by the LDA independently, matching the job titles and definitions set by this study. The reliability of the results averaged 84.7% for the five coders, confirming the significance of our results.

As with the job applications, three international job classification standards (ISCO, SOC, and O*Net) were used to establish the job occupations' titles and definitions for each group of the extracted jobs. To verify the validity of the titles and definitions of the derived nine job occupations, an intercoder reliability analysis between the five experts in the SW field was performed. As with the job cluster development, each of the five experts matched each of the set titles and definitions for independently separated clusters of job occupations. The matching results were found to be appropriate based on the consistency of the results of the experts, demonstrating significance for the developed job occupation system, showing an average of 82.2% reliability in the results by the group of five experts and the results of the job occupation system developed in this study.

1) http://www.ilo.org/public/english/bureau/stat/isco/

2) https://www.bls.gov/soc/, 2010 financial edition is the latest standard.

3) https://www.bls.gov/soc/, It has been available on the website since 2016.

<Table 2> The Result of LDA for Job and Job Occupations

	Job Occupation 1		Job Occupation 2		Job Occupation 3	Job Occupation 4			Job Occupation 5	Job Occupation 6		Job Occupation 7	Job Occupation 8	Job Occupation 9	
Job Cluster	Job 1	Job 2	Job 3	Job 4	Job 5	Job 6	Job 7	Job 8	Job 9	Job 10	Job 11	Job 12	Job 13	Job 14	Job 15
Keyword (co-occurrence)	Management (0.2406)	Design (0.1862)	Design (0.259)	Design (0.259)	Engineer (0.1336)	Development (0.0192)	Development (0.1327)	Development (0.1302)	Hardware (0.1267)	Web (0.109)	IT (0.1466)	Development (0.1263)	Marketing (0.1693)	Development (0.2862)	Development (0.3717)
	ShoppingMall (0.0832)	ShoppingMall (0.1256)	Web (0.2222)	System (0.0632)	Network (0.1117)	Web (0.0761)	Program (0.1293)	Operation (0.1286)	Aided Design (0.1154)	Editing (0.0913)	Development (0.0891)	Web (0.0875)	Online (0.111)	java (0.1262)	Web (0.1351)
	Product (0.0675)	Web (0.0701)	Management (0.0564)	Data Processing (0.0627)	Security (0.0899)	c (0.0703)	Server (0.1225)	Software (0.0938)	Design (0.0696)	Planning (0.0754)	Project (0.0759)	Mobile (0.0852)	Design (0.0869)	Operation (0.0483)	android (0.0709)
	Homepage (0.058)	Openmarket (0.0667)	Photoshop (0.048)	Advertising (0.0577)	Installation (0.0854)	c++ (0.0454)	Computer (0.0837)	Management (0.0609)	Development (0.0687)	Production (0.0726)	java (0.0699)	Design (0.0473)	Planning (0.0826)	System (0.0374)	App (0.0687)
	Online (0.0556)	3d (0.0543)	ShoppingMall (0.0398)	IT (0.0369)	System (0.0557)	Design (0.0426)	System (0.0452)	Analysis (0.0445)	Support (0.0473)	Broadcasting (0.0515)	Viral (0.0515)	Database (0.0423)	cs (0.053)	jsp (0.0274)	ios (0.0499)
	Design (0.0505)	Planning (0.0505)	Installation (0.0286)	Web (0.0344)	Equipment (0.0506)	Image (0.0395)	linux (0.0442)	Planning (0.0421)	Coltrol (0.036)	PM (0.0437)	Web (0.0398)	Planning (0.0345)	UI (0.0462)	Aided Design (0.025)	php (0.041)
	Production (0.042)	Online (0.0321)	Internet (0.0276)	Production (0.03)	Management (0.0497)	Contents (0.0321)	Engineer (0.0368)	Service (0.0356)	Graphic (0.0355)	Education (0.0404)	Social (0.0373)	Management (0.0301)	UX (0.0422)	c (0.0217)	Mobile (0.0271)
	Page (0.0363)	Operation (0.0254)	Homepage (0.025)	cg (0.0294)	Information (0.0474)	Game (0.0266)	Web (0.0333)	Graphic (0.0287)	Web (0.0313)	Cording (0.0391)	Marketing (0.037)	Production (0.0232)	ShoppingMall (0.0411)	Research (0.0212)	java (0.0243)
	Blog (0.0245)	Contents (0.022)	Planning (0.0237)	Planning (0.0241)	Building (0.0265)	html5 (0.0263)	Proposal (0.0216)	Center (0.0278)	Machine (0.0301)	Mobile (0.0386)	OpenMarket (0.0346)	Operation (0.0231)	Digital (0.0307)	Database (0.02)	Planning (0.0177)
	Planning (0.0243)	Production (0.0186)	Operation (0.0153)	Editing (0.016)	Development (0.0227)	ShoppingMall (0.0257)	Network (0.0206)	Program (0.0268)	Electronic (0.0296)	Selling (0.0328)	PC (0.0338)	Project (0.0219)	Web (0.0293)	Web (0.0196)	Operation (0.0174)

Finally, the job classification system based on data analysis was developed to reflect the demand in the SW industry by separating fifteen jobs from a total of nine job occupations, as shown in <Table 3>. Each job was compiled by referring to the definitions of the international job classification standards and developed by system design analysts, content creators, and human-computer interaction (HCI) consultants of the newly developed job occupations. In addition, data-based classification systems were developed based on the demand of the SW industry as detailed jobs such as media content creators, social media

<Table 3> The Classification and Definition of Job and Job Occupations for SW Industry

<table>
<tr><th>Job occupation</th><th colspan="2">Job title and explanations</th></tr>
<tr><td rowspan="3">Job Occupation 1</td><td colspan="2">Web Administrators and Planners
: planning and managing contents, design, and services required to operate online channels such as online shopping malls, open markets, blogs and web pages</td></tr>
<tr><td>Job 1</td><td>Web Administrator
: the management of bulletin boards on corporate websites and the various issues that arise from websites</td></tr>
<tr><td>Job 2</td><td>Web Planner
: Specialist in charge of planning and marketing related to website construction</td></tr>
<tr><td rowspan="3">Job Occupation 2</td><td colspan="2">Digital Content Designers
: designing elements such as online advertising, computer graphics, product design, etc. on a variety of web</td></tr>
<tr><td>Job 3</td><td>Web Content Designer
: to design web formats for web-based graphics, audio, and video-related elements using SW</td></tr>
<tr><td>Job 4</td><td>Graphic Designer
: designing and creating logos, graphics, etc. for specific trading and promotion activities</td></tr>
<tr><td rowspan="2">Job Occupation 3</td><td colspan="2">Computer Network Administrator
: conducting research and analysis for the structure and implementation of a Fitter network, operates a computer network architecture design, and tests an implemented network.</td></tr>
<tr><td>Job 5</td><td>Computer Network Administrator
: Configuration status of the entire network system and monitor the operation of the system in real time.</td></tr>
<tr><td rowspan="4">Job Occupation 4</td><td colspan="2">System and Software Developers
: analyzing and evaluating systems, web pages, SWs, or existing programs for new servers to identify users' needs and to plan, develop, test, and operate system SW, web pages, SW, and applications to meet users' needs.</td></tr>
<tr><td>Job 6</td><td>Software Developer
: developing various SWs, such as office programs, accounting programs, database tools, and statistical programs, and changing the SW environment according to the use environment of the computer system</td></tr>
<tr><td>Job 7</td><td>System Developer
: developing programs related to R&D, design and system development of System S/W</td></tr>
<tr><td>Job 8</td><td>Software Analyst
: planning, designing, evaluating and experimenting with SW to achieve the purpose of providing the SW's design, preparation, experimentation, and individualized performance and for character applications</td></tr>
</table>

<Table 3> The Classification and Definition of Job and Job Occupations for SW Industry(Cont.)

Job occupation	Job title and explanations	
Job Occupation 5	System Analyst* : providing planning, launching and managing IT technology-related projects and unified system management of corporate management and technical issues	
	Job 9	System Analyst : overall management to design and analyze systems and review system capacity, development work procedures and schedules
Job Occupation 6	Content Creators* : developing marketing content for platform content as a form of oral marketing and marketing campaigns	
	Job 10	Media Content Creator* : project manager to plan and produce the content on a web page, broadcast, or mobile page, and the overall process of planning, creating, and editing the content
	Job 11	Social Media Manager* : carrying out social marketing that creates oral effects of consumers as part of marketing activities such as web pages and open markets
Job Occupation 7	Database Administrator : professional design, development, operation, maintenance and support for optimal performance and security of information systems	
	Job 12	Database Administrator : analyzing and managing and tuning database systems.
Job Occupation 8	HCI Consultant* : developing and consulting user friendly and accessible computer systems and websites based on human computer interaction (HIC).	
	Job 13	HCI Consultant : Designing usability and user experience, based on the user interface, interaction design
Job Occupation 9	Applications Developers : setting the scope and goals of development of an application SW under a comprehensive review and analysis of its capabilities, performance requirements, and operating environment (OS, network, user interface, etc.)	
	Job 14	Software Applications Developer : coding by a programming language, such as java, c++, managing and developing applications and application
	Job 15	Mobile Applications Developer* : coding and managing mobile programming languages such as Java and PHP : developing and operating apps or mobile webpages for Android and iOS

managers, mobile developers, HCI consultants, and system design analysts were newly created.

This classification system is a classification system analyzed in accordance with the employment announcement data of the SW industry. We can reduce instability caused by the inconsistency between the existing Worknet and NCS (National Competency Standards) classification system by classifying job occupations and jobs that actively reflect industrial characteristics.

In addition, this study identifies the simultaneous listing rate in the documents of key words within each category (the same frequency of the light source), with significant implications regarding information on the skills required or of high importance and areas of interest.

Nevertheless, this job classification system may change the classification of job occupations and jobs over time with additional jobs and job posting data. Therefore, the current classification system of jobs and job occupations should undergo the re-classification process within six months to one year to establish a new classification system. It is also necessary to consider the need for reacquiring the word co-occurrence of keywords.

Ⅴ. Conclusions

In this study, the job and job occupation classification system was developed by applying big data analysis. The existing Worknet and NCS classification systems were developed to further fulfill the unclear classification criteria. Accordingly, we developed a framework to further satisfy the unclear classification criteria of the Worknet and NCS classification systems.

As a result of the analysis, we verified that the Worknet and NCS job classification systems are inconsistent, with non-mutually matching subdivided jobs between them. Second, to resolve the inconsistency between the Worknet and NCS systems, we analyzed the contents of the Worknet job announcement of the SW industry using text mining techniques. We gathered the total job announcements from Worknet, and then derived a job classification system based on LDA topic modeling. Thus, fifteen jobs were classified according to nine job families, with three of them being new job families and five being new jobs. Third, we defined each occupation, and divided their tasks by referring to the three international job standards for each occupation classification (ISCO, SOC, and O*Net).

This study has practically developed a classification system to improve the Worknet and NCS classification systems. First, from a practical viewpoint, we can provide guidelines for formulating a training policy for SW specialists through the demand for workforce in the SW industry, since it reflects the job classification system by utilizing the actual job postings of companies in the SW field. Second, we can identify keywords in job announcements by each job category in the newly developed job classification system, and better recognize job competency.

Since SW is one of the rapidly changing industries, big data-based analysis, which analyzes the job announcements by word-level, can be used as an optimized tool to identify the characteristics of the rapidly changing SW industry in the future. Using the job classification system developed in this study, we can objectively evaluate the competence of each job that is classified and identify the changes in industry demand timely, if the data are continuously collected. This will be useful not only for complementing the existing NCS system, but also for implementing timely policies in response to changes in the demand of the SW industry. Similarly, we expect the results of this paper to narrow the gap between demand and supply in the SW field (i.e., the SW industry and higher education system) by improving the academic curricula.

The limitations of this study and the future direction of research are as follows. First, it is necessary to collect and analyze disaggregated data on multitask recruitment announcements. In the data analyzed in this study (20,408 cases), approximately 156 announcements were found to have been made. Since the number of multiple announcements in this study was very small, being 0.7% of the total number of job postings, it did not significantly affect the data analysis-based job and job classification system, which was included in the analysis. Consequently, multi-employer announcements can result in a mix of keywords for different tasks, if the number is greater than 50% of the total public offering. Therefore, it is necessary to develop and improve methods to reflect future multi-employer announcements into the classification system through separate natural language processing and data-refining techniques. Second, in this study, data on one-word units such as 'Web', 'Design', and 'Develop' were extracted to analyze probability and frequency of appearance. Future research should conduct analysis based on more advanced task data by utilizing a method for refining the pairing of two words such as 'Web-Design', 'System-Design', 'Web-Development', and 'App-Development'. Finally, based on the data analysis of the SW job and occupational classification system developed in this study, SW-related education systems in universities can solve the problem of inter-disciplinary gaps with respect to the SW workforce.

Acknowledgements

This paper was (partially) supported by the New Professor Research Program of KOREATECH in 2018.

This research was supported by the Soonchunhyang University Research Fund.

<References>

[1] Apte, U. M., Karmarkar, U. S. and Nath, H. K. (2008). Information services in the U.S. economy: Value, jobs, and management implications. *California Management Review, 50*(3), 12-30.

[2] Arntz, M., Gregory, T., and Ulrich, Z. (2016). *The risk of automation for jobs in OECD countries.*

[3] Bakhshi, H., Downing, J. M., Osborne, M. A., and Schneider, P. (2017). *The future of skills: Employment in 2030.* London: Pearson.

[4] Chen, H., Chiang, R. H. L., and Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly, 36*(4), 1165-1188.

[5] Cho, W. Y., and Lee, D. H. (2016). *The gold vein of future jobs, Software.* Software Policy & Research Institute.

[6] Doan, A., Ramakrishnan, R., and Halevy, A. Y. (2011). Crowdsourcing systems on the World-Wide Web. *Communications of the ACM, 54*(4), 86-96.

[7] Dudley, N. M., Orvis, K. A., Lebiecki, J. E. and Cortina, J. M. (2006). A meta-analytic investigation of conscientiousness in the prediction of job performance: Examining the intercorrelations and the incremental validity of narrow traits. *Journal of Applied Psychology, 91*(1), 40-57.

[8] Frey, C. B. and Osborne, M. (2013). *The future of employment.*

[9] Frey, C. B., Osborne, M., Holmes, C., Rahbari, E., Garlick, R., Friedlander, G., and Chalif, P. (2016). *Technology at work v2.0: The future is not what it used to be.* CityGroup and University of Oxford.

[10] Gardner, T. M. (2005). Interfirm competition for human resources: Evidence from the software industry. *Academy of Management Journal, 48*(2), 237-256.

[11] Igbaria, M., Greenhaus, J. H. and Parasuraman, S. (1991). Career orientations of MIS employees: An empirical analysis. *MIS Quarterly, 15*(2), 151-169.

[12] Kim, J. J., Kang, E. Y., Kim, D. K., Park, G. Y., Lee, R., Lee, Y. S., and Jang, J. H. (2016). *New job research through domestic and overseas job comparison analysis.* Korea Employment Information Service.

[13] Lee, D. H. (2015). *A study on the future job in SW oriented society: The threat and countermeasures of computerization*. Software Policy & Research Institute.

[14] Manyika, J. (2017). A future that works: AI automation employment and productivity. *McKinsey Global Institute Research*, Tech. Rep.

[15] Maybury, M. T. (2004). *New directions in question answering*. Cambridge, MA: The MIT Press.

[16] O'Reilly, T. (2005). *What Is web 2.0? Design patterns and business models for the next generation of software*. September 30, Retrieved from http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html.

[17] Pang, B., and Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval, 2*(1/2), 1-135.

[18] U.S. Government (2016). Artificial intelligence, automation, and the economy. *Executive Office of the President*, 18-19.

[19] Vogler-Ludwig, K., and Dull, N. (2014). *The German labour market in the Year 2030*. German Ministry for Labour and Social Affairs.

[20] World Economic Forum (2016). The future of jobs: Employment, skills and workforce strategy for the fourth industrial revolution. *Global Challenge Insight Report, World Economic Forum*, Geneva.

◆ About the Authors ◆

Ho Lee

Ho Lee is an assistant professor in the Department of Future Technology at Korea University of Technology and Education. He completed a Ph.D. in Information Systems at Yonsei University, Korea and received his Bachelor of Science in Computer Science from State University of New York at Sony Brook, USA. His current research interests are in the areas of anonymity, online behavior, knowledge management, job change and data analytics.

Jaewon Choi

Jaewon Choi is an assistant professor of Business Administration, Global Business School, Soonchunhyang University. His research areas are investigating big data analysis, social network analysis, block chain, personalized intelligent agents in e-commerce and m-commerce. He published papers on Journal of Electronic Commerce Research, International Journal of Electronic Commerce, Cyberpsychology Behavior and Social Networking, and other journals.

Submitted: September 25, 2019; 1st Revision: October 22, 2019; Accepted: October 29, 2019