A Study on the Analysis of IT-related Occupational Cluster using Big Data

Do-Yeon Kim, Seong-Taek Park, and Mi-Hyun Ko

Abstract—**Many countries and firms are exploiting and utilizing big data for a solution of social problems and creation of jobs, and economic effects. This study is tried to understand trend of IT-related occupational fields by systematic analysis of data and analyzed the most posted IT-related jobs by classification standards. We used Java programming language to analyze IT-related occupational fields and collected IT job data posted on online recruitment web site "Saramin (http://www.saramin.co.kr)". We constructed database by selecting 17,781 refined data. We divided IT-related occupational fields into '1. Developer (Programmer)', '2. Technician', '3. Planner', '4. Designer', '5. Counselor', '6. Operator', '7. etc.' and analyzed frequency of jobs posted on online recruitment web site by each field. As results, the most appeared job related to "developer" was software developer. Appearance frequency of web designer was highest among designer jobs. When it comes to "planner", the most appeared thing was online marketing. In case of "operator" and "counselor", those were network administrator and network solution sales respectively. Among "technician" fields, system maintenance was most frequently appeared. Finally, the most appeared job related to "etc." fields was computing job. We introduced crawling work in order to examine IT-related jobs trends. Also, we classified IT-related jobs into 7 fields. Researcher who wants to understand job trend of various fields can use our method and process. Also, we think that the field of vocation research can get a very in depth look from our study.**

Index Terms—**IT-related jobs, Big Data, Java, R programming, Online recruitment web site**

I. INTRODUCTION

In current uncertain environment, we have a desire to make sure of uncertainty [1]. Along with the arrival of the era of big data, many big data analysis methods have been applied to the ontology similarity calculation and ontology mapping [2]. Recently, data is created and consumed dramatically by diffusion of smart device and mobile Internet environment and emergence of social media with the development of information technology (IT). As type of data has diversified, processing technology of it becomes the key issue of ICT (Information Communication Technology) fields [3]. Today's Information Technology applications rely on various computer technology such as cloud computing [4].

Manuscript received June 17, 2017, revised August 20, 2017

D.Y. Kim is with Center for Free Semester and Career Experience Program, Korea Research Institute for Vocational Education & Training, Republic of Korea. (e-mail: lloook91@naver.com)

S.T. Park is with Dept. of Management Information Systems, Chungbuk National University, Republic of Korea. (e-mail: solpherd@chungbuk.ac.kr)

M.H. Ko is with Div. of Policy Research, Korea Institute of Science and Technology Information, Republic of Korea. (corresponding author e-mail: mihyungo@kisti.re.kr)

Thus, the importance and value of Big Data are receiving attention to the public. We expect fields like politics, medicals, education, and business will utilize Big Data strategically through the process of analysis and screening of data. Republic of Korea has a lot of concern for the field of job and employment this study deal with and the related subjects are various and segmented. Also, change in its policy is too frequent. The fields of job and employment are creating much data, but analysis system of atypical data on online recruitment web site posted is not doing enough.

As discussed above, the big data era when data can be economic asset has just come [5]. Many countries and firms expected that big data decide success or failure of them and create source of forthcoming economic value. McKinsey, The Economist, and Gartner researched prediction of market fluctuation utilizing big data and cases and effects of economic value creation such as discovery of new markets [6][7]. Especially, many countries and firms are exploiting and utilizing big data for a solution of social problems and creation of jobs, and economic effects. Big data analysis methods are very diverse principal component analysis (PCA), singular value decomposition (SVD) and hierarchical clustering, just can be applied with complete datasets without missing values [8]. Recently Online Social Networks (OSNs) have been studied extensively [9]. Especially, data increases, machine learning research has increased [10]

This study is tried to understand trends of IT-related occupational jobs by systematic analysis of data and analyzed the most posted IT-related occupational jobs by classification standards

II. LITERATURE REVIEW

A. The concept and characteristics of big data

The national informatization policies committee (2011) defined that big data is the technology to extract valuable information by utilizing and analyzing massive data and to actively response and predict changes over the created knowledge [11]. According to Sajana et al. (2016), Big Data are the large amount of data being processed by the data mining environment [12]. Noh and Lee (2015) said that big data platform is software that uses computing power based on distributed parallel method in order to process large amounts of data [13]. Also, Samsung Economic Research Institute (2010) defined big data as huge data sets that cannot be handled by existing enterprises management and analysis

system. But it seems that big data are not simple linear meaning. IT consulting group 'Gartner' explains the concept of big data with so called '3V (Volume, Velocity, Variety)' and '1C (Complexity)' [14]. 'Volume' means enormous data size unlike existing data paradigm. 'Velocity' means that data changes in real time and 'Variety' implies that big data include semi-structured data in addition to structured data [15]. Russom (2011) emphasized the importance of network to treat big data. Extant literatures in aspects of diversity of big data have studied big data around the form of various data, but it needs research of channels to collect data [16].

B. The concept and characteristics of big data

Because interest of firms and governments in big data is increasing, studying affecting factors from introduction of big data is very important. Especially, many firms and institutes lacked the preparation of introduction of big data. So, we thought that research to explore affecting factors for the successful introduction of big data is timely [17]. In the future, value creation by rapid decision making through Big Data analysis in the field such as public administration, finance, distribution, medical, manufacturing gradually will increase [14]. All fields have to overcome many obstacles in order to extract value from Big Data, but by fields, it is not easy to overcome the many difficulties structurally. Especially, public service such as education because of low willingness to use data and shortage of available data has many obstacles. The easiness degree of acquisition of potential value by fields can be tabulated as follows (See Table 1) [4].

III. PROPOSED WORK

We used Java programming language to analyze IT-related occupational field and collected IT job data on online recruitment web site "Saramin" (http://www.saramin.co.kr) posted. Data was collected between 1 January 2012 and 17 March 2016 and total amount of data collected are 403,112. From among these, we subtracted only IT-related occupational cluster through deletion of error data and constructed DB (database) by selecting 17,781 refined data. After that, we classified data by occupational cluster and confirmed the most appeared IT jobs by using R package (Shown in Figure 1).

TABLE I
THE EASINESS LEVEL OF ACQUISITION OF POTENTIAL VALUE

| Category	Fields	Professional manpower	Holding of IT asset	Will to use data	Data availability	AVG.
Commodity	Manufacture	5	2	3	5	5
	Service	2	3	3	3	3
	Public sector	1	3	5	5	4
	Computer, Electronics	4	3	3	3	4
	Real estate,	1	5	5	1	3
	Wholesaling	3	4	3	1	3
	Information	4	5	2	4	5
Service	Transportation, warehousing	3	5	1	4	4
	Retail	2	2	2	3	2
	Waste management, purification	3	2	4	2	3
	Lodging,	1	1	4	4	2
	Other services excluding public administration	2	2	3	2	1
	Arts, entertainment, leisure	1	4	1	3	1
	Finance,	5	2	2	4	4
	Science, technology service	5	4	2	2	3
	Enterprise management	5	5	1	1	2
Public	Government	-	-	1	2	1
	Education	4	1	1	1	1
	Health, social security	4	1	5	5	4
	public services(electricity, water, and so on)	3	4	5	5	5
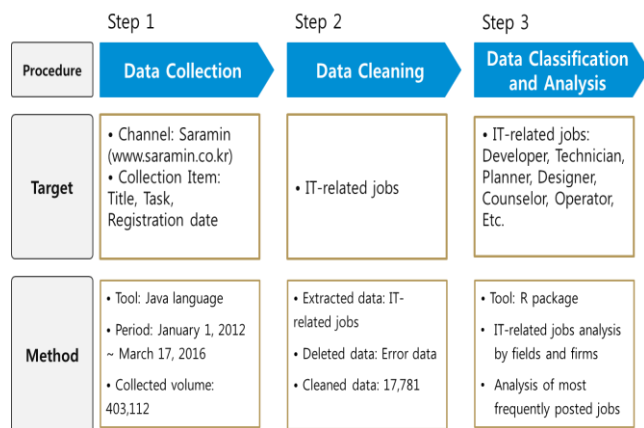
Fig. 1. Analysis Process

Figure 2 explains collecting method we used. As you can see, it is a console screen that represents Java programming code and crawling result to collect postings of IT-related job in Saramin web site. We excluded unnecessary words such as "recruitment (chae-yong in Korean)", "employment notice (chae-yong-gong-go in Korean)", "help-wanted (goo-ham-ni-da in Korean)" from crawling works. Data lists of IT-related jobs gained from Saramin were broken into following clusters: (1) Web planning, (2) Web design, (3) Web programmer, (4) System programmer, (5) Network security, (6) HW/SW, (7) IT/Design training, (8) Game, (9) Contents site, (10) Webmaster, (11) ERP, (12) System analysis, and so on. But we made a new classification, because these clusters have no standard.
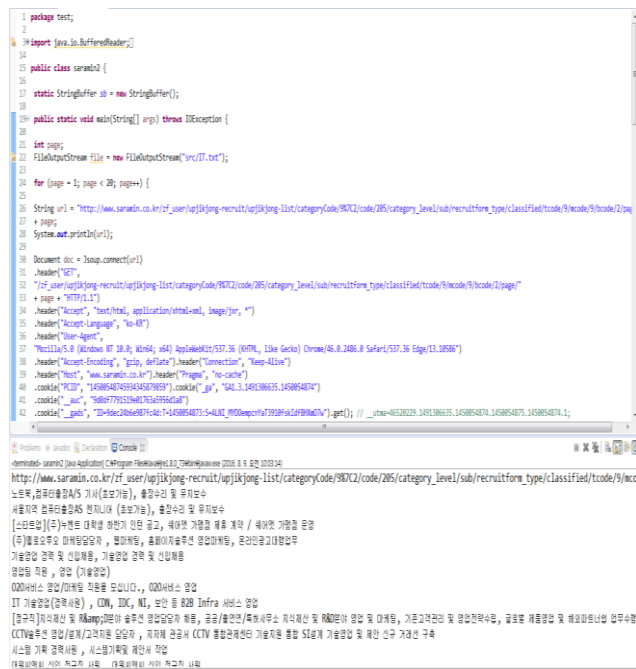

Fig. 2. Crawling work and result screen using Java language

IV. Conclusion

As data size collected by type of job (IT·Internet, Media, specialized job, education) was too vast, we have no choice but to divide IT-related jobs into '1. Developer (Programmer)', '2. Technician', '3. Planner', '4. Designer', '5. Counselor', '6. Operator', '7. etc.' (See Table 2).

TABLE Ⅱ
CLASSIFICATION OF IT-RELATED JOBS AND ITS EXAMPLES

No.	IT-related jobs	Examples
1	Developer	Programmer, Coding, web-page production
2	Technician	Engineer, maintenance
3	Planner	Marketing, MD
4	Designer	Photoshop, Illustrator, UI, UX
5	Counselor	Advisor, Sales work
6	Operator	Administrator
7	etc.	IT instructor, Computing job, Analyst, translator, security manager

Table 3 shows collection of IT-related jobs by type of job. Designer has the largest amount of posts in the fields of IT·Internet, followed by developer and planner. Also, designer has the largest amount of posts in the fields of media once again, followed planner and operator. In the field of education, 'etc.' has the largest amount of posts while developer is the largest in the fields of specialized jobs, followed planner and operator. 'etc.' recorded the largest posts in the field of education because it included IT lecturer of diverse domain (See Table 3).

TABLE Ⅲ
THE NUMBER OF COLLECTION OF IT-RELATED JOBS BY THE FIELDS OF JOBS
DATE COLLECTED: MARCH 17, 2016 19:00 / (UNIT: POST)

No,	IT-related jobs	The field of jobs				Total
		IT·Internet	Media	Specialized jobs	Education	
1	Developer	3,605	24	167	22	3,818
2	Technician	1,114	21	62	13	1,210
3	Planner	2,400	259	76	10	2,745
4	Designer	3,214	666	74	28	3,982
5	Counselor	912	2	12	4	930
6	Operator	2,065	36	30	7	2,138
7	etc.	507	4	43	274	828
	Total	13,817	1,012	464	358	15,651

We used R package in order to analyze the largest posted job from data above. Table 4 ranked from No.1 to No.30 by frequency of IT-related jobs. Web designer (2,575) is the largest in the frequency, followed by software developer (2,361), online marketer (1,138), and technician (800) (See Table 4).

TABLE IV
THE RANK OF IT-RELATED JOBS BASED ON FREQUENCY

(UNIT: POST)

Rank	Jobs	Frequency
1	Web designer	2,575
2	Software developer	2,361
3	Online marketer	1,138
4	Technician	800
5	Programmer	793
6	Solution sales	757
7	System maintenance	694
8	Web planner	687
9	Operator	647
10	Network administrator	553
11	Computing jobs	476
12	Network engineer	455
13	Web developer	444
14	Shopping mall operator	439
15	Homepage production	271
16	Customer Counselor	254
17	Game planner	249
18	Server administrator	214
19	Photoshop	211
20	Online sales	207
21	Mobile application developer	181
22	UI, UX Designer	165
23	Sales management	157
24	Graphic design	137
25	Viral marketing	131
26	Data analyst	131
27	IT Instructor	111
28	Security manager	110
29	Game developer	100
30	Illustrator	75

The top 10 most appeared by occupational field subtracted are as follows. The most appeared job related to "developer" was software developer (2,361), followed by programmer (793), web developer (444), and computer game developer (100). Next, the most appeared job related to "designer" was web designer (2,575), followed by Photoshop (211), UX/UI designer (165), graphic designer (137), and illustrator designer (75). And the most appeared job related to "planner" was online marketing (1,138), followed by web planner (687), game planner (249), viral marketing (131), product planner (45), and web marketing (45). Also, the most appeared job related to "operator" was network administrator (553), followed by shopping mall administrator (439), server administrator (214), sales manager (157), production manager (49), and blog administrator (46). The most appeared job related to "counselor" was network solution sales (757), followed by online sales (207), CS (Customer Service) counselor (73), technology counselor (67), mobile counselor (49), and sales MD (Merchandiser)(47). The most appeared job related to "technician" was system maintenance (594), followed by software engineer (98), network engineer (85), Internet setup and maintenance (76), system engineer (31). Finally, the most appeared job related to other was computing job (530), followed by data analysis specialist (147), OA instructor (127), security specialist (110), programming language instructor. When it comes to IT instructor, OA instructor is the most frequently appeared job, followed by programming instructor, web instructor. In addition, we found that job called by IT translator is newly-listed.

Frankly speaking, we tried to track IT-related job trends. Come to think of it, our application method was right on targets. When it comes to implication of our study, first, we classified IT-related jobs into seven fields. Next, we collected the number of postings of IT-related jobs with crawling works and analyzed frequency of its. Researcher who wants to understand job trend of various fields can use our method and process. Especially, we think that the field of vocation research can get a very in depth look from our study. Like Google predicted flu with search recording of users a long time ago, these works can forecast job trend in any field.

REFERENCES

[1] S. Sang, "Pricing and Service Decisions of Supply Chain in an Uncertain Environment," Engineering Letters, vol. 24, no. 4, pp 441-448, 2016
[2] W. Gao, L.L. Zhu, and Y. Guo, "Multi-dividing infinite push ontology algorithm," Engineering Letters, vol. 23, no.3, pp 132-139, 2015
[3] J. Gantz, and D. Reinsel, "Extracting Value from Chaos," IDC iview, pp 1-12, 2011
[4] Z. Wang, J. Wang, B. Li, Y. Liu, and J. Ma, "Online Cloud Provider Selection for QoS-Sensitive Users: Learning with Competition," IAENG International Journal of Computer Science, vol. 43, no.3, pp 310-317, 2016
[5] S. LaValle, E. Lesser, R. Shockley, M.S. Hopkins, and N. Kruschwitz, "Big data, analytics and the path from insights to value," MIT Sloan Management Review, vol. 52, no. 2, pp 21-32, 2011
[6] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A.H. Byers, "Big data: The Next Frontier for Innovation, Competition, and Productivity," McKinsey Global Institute., 2011

[7] A. Gandomi. and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," International Journal of Information Management, vol. 35, no. 2, pp 137-144, 2015

[8] Y. He, and D.C. Pi, "Improving KNN Method Based on Reduced Relational Grade for Microarray Missing Values Imputation," IAENG International Journal of Computer Science, vol. 43, no. 3, pp 356-362, 2016

[9] E.Y. Baagyere, Z. Qin, H. Xiong, and Q. Zhiguang, "The Structural Properties of Online Social Networks and their Application Areas," IAENG International Journal of Computer Science, vol. 43, no. 2, pp 156-166, 2016

[10] M. Langovoy, "Machine learning and statistical analysis for BRDF data from computer graphics and multidimensional reflectometry," IAENG International Journal of Computer Science, vol. 42, no. 1, pp 23-30, 2015

[11] G. Lee, and Y. H. Kwak, "Open government implementation model: a stage model for achieving increased public engagement," In Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times, ACM, pp 254-261, 2011

[12] T. Sajana, C.M. Sheela Rani, and K.V. Narayana, "A Survey on Clustering Techniques for Big Data Mining," Indian Journal of Science & Technology, vol. 9, no. 3, pp 1-12, 2016

[13] K.S. Noh, D.S. Lee, "Bigdata Platform Design and Implementation Model," Indian Journal of Science & Technology, vol. 8, no. 18. pp 1-8, 2015

[14] W. Fan, and A. Bifet, "Mining big data: current status, and forecast to the future," ACM SIGKDD Explorations Newsletter, vol. 14, no.2, pp 1-5, 2013

[15] J. Wareham, A. Rai, and G. Pickering, "Standardization in Vertical Industries: An Institutional Analysis of XML-Based Standards Infusion in Electricity Markets. Electronic Markets," vol. 15, no. 4, pp 323-334, 2005

[16] P. Russom, "Big data analytics," TDWI best practices report, fourth quarter, 2011.

[17] A. McAfee, E. Brynjolfsson, and T.H. Davenport, "Big data: the management revolution, Harvard Business Review," vol. 90, no. 10, pp 60-68, 2012