

Analysis on Big Data by Performance Factors of Creative Education using Semi-structured Data-based Twitter

Kil-Hong Joo¹, Ji-Hoon Seo² and Nam Hun Park^{3*}

Dept. of Computer Education, Gyeongin National University of Education, Korea

Dept. of Computer Science Engineering Incheon University, Korea

Dept. of Computer Science, Anyang University, Korea

*khjoo@ginue.ac.kr, sserz@inu.ac.kr, *nmhnpark@anyang.ac.kr*

Abstract

As various forms of big data, which includes but not limited to, large volume texts, voice data and videos, are being accumulated whilst the waves of the information age are accelerating progressively, the number of inter-disciplinary analysis solutions with capabilities to use such information is increasing, and accordingly, the developments, such as the drop of costs required for data storage and various Social Network Services, have brought forth the quantitative and qualitative stretch of the data. The phenomenon makes it possible to achieve the types of data usage which were not available in the past, and thus the potential values and leverage of data are on the rise. Studies that that apply such inter-disciplinary analysis system for the improvement of the educational system to suggest future-oriented education system are being carried out at progressive levels. This study has carried out an analysis on big data with Twitter as its subject and suggested, via the natural language process of data and frequency analysis, the quantitative scale indicative of how various issues and performances relating to creative education in South Korea have been handled.

Keywords: *Creative education, big data, pattern algorithm, association role*

1. Introduction

As most public institutions of South Korea and its departments have initiated to provide and be equipped with effective systems for information management & analysis and information use in agreement with government 3.0, higher demands for systematic analyses and applications of big data are being made to seek education performance analyses and improvement proposals. The use of digital devices and social networking services has become widespread with the development of informatization, and, as a result, the amount of data being accumulated is increasing. The phenomenon not only enables us to make use of the data, which was not attempted previously, and is enlarging the potential values and leverage of the data. There exist many fields in which these aspects can be applied, and measures apt for carrying out appropriate education can be suggested, if the educational data previously accumulated are properly utilized. Methods that have previously existed were consisted of brainstorming, Delphi, panel of experts and et cetera, whereas, in this study, results of analysis that are based on quantitative data can be suggested by analyzing the big data objectively, for it uses Twitter via online mediums. Therefore, this study, with creative education, one of the country's policies that is being enforced and highlighted, as its keyword, has made discoveries and carried out extensive researches on issues and performances creative education have achieved, if any, by collecting the South Korean Twitter data consisted of data of the past to the present time.

* Corresponding Author

2. Related Work

Studies that cover big data of educational information include researches carried out by Jaeyeong Choi (2012), Yongsang Jo *et al.* (2013) and Wooju Kim (2013). Jaeyeong Choi (2012), in his study, has discussed big data in the Smart Education environment. The study estimates the types and sizes of educational data which can occur within the scope of the Smart Education environment and suggests methods for utilizing big data in the Smart Education environment. By having estimated the educational big data that occurs within the scope of the Smart Education environment, it has estimated that seven million elementary, middle and high school students throughout the county produce 7.0 terabytes of data a day and the sum data per year amounts to 2.56 petabytes. In addition, the study has suggested customized learning support, contents development, improvement of study courses and educational policies, and student guidance as methods for utilizing the educational big data.

The study, by reinterpreting the examples of technical applications in the learning analysis suggested in the policy brief of UNESCO IITE (Institute for Information Technologies in Education) that becomes compatible with the South Korean environment, suggests the possibilities of technical applications of the learning analysis expected as in the following.

First is the learning platform's LMS/VLE Analytics Dashboard. The LMS/VLE Analytics Dashboard, which has been added to the learning platform, is a visualized function that provides the log data to enable individual users or groups to have a more transparent understanding, and there has been a case study at the Arizona State University through which all data available for collection were comprehensively analyzed to get a better understanding of the individual student.

Second is the predictive analytics. Predictive analytics, as a method that analyzes users by using the patterns of users' statistical information, has the advantage of being able to suggest predictive measures at a time desired, which is achieved by analyzing the patterns of the users who are at dangerous or outstanding levels. The Course Signal of Purdue University is one of the works related to this subject.

Third is the Adaptive Learning Analytics. It is a model that provides feedback on a particular subject by measuring the learner's level of understanding and relevant digital resources to the learners. The courses offered by Open Learning Initiative, which were created based on the research findings of Carnegie Mellon University, are one of its related works.

Fourth is the Social Network Analytics. Analysis on the impacts of the learner's personal relationship settings and group's structure on learning in the field of education is required.

Fifth is the Discourse Analytics. Discourse analysis is a model designed for determining the internal stability of the activities and is used for analyzing various factors, including how a teacher corrects essays or debate speeches, how writing and speaking affect learning and how computers recognize an excellent argument.

Sixth is the assessment using Itching the study of Wooju Kim (2013), actual analysis on big data has been carried out by using publicly disclosed educational information and access log data. In this thesis, by exploring the interests of the users by regional groups and schools by analyzing the publicly disclosed educational information and access log data, which are standard and semi-structured data, interests of the users have been discovered.

3. Proposed Method

In this thesis, a five-phase process, which includes the collection of semi-structured data, securement and classification of historic data through collection, data preprocessing, natural language processing of the Korean grammar and big data analysis, had been carried out via Twitter as a process for the analyzation of performance factors.

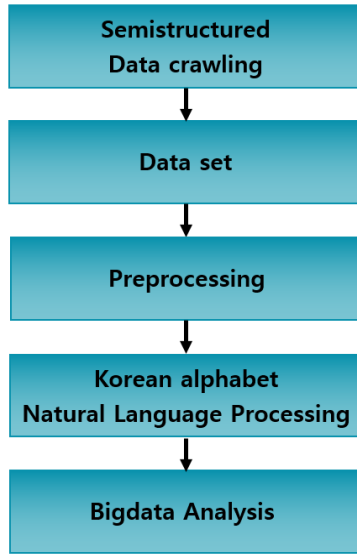


Figure 1. Process of Data Analysis

3.1. Collecting Semi-structured Data for Big Data Analysis

Collecting historic data is vital for carrying out a quantitative analysis on the recent analysis of creative education of South Korea. Historic data may have subtle effects on the results of big data analysis based on the amount, cycle and route pursuant to the collection, and the errors of the data being resulted can be applied in a significant way. In this thesis, online media was used as the subject for the collection of historic data and Twitter was selected as the data type. For the data cycle, documents about creative education published in between year 2013 and 2015 through online media were gathered.

G1	A	B	C	D	E	F	G
1	181245986134155294	2014-01-01 00:03:08	Caria_Rutz_bot	Caria_Rutz_bot	ko	0	올림픽 개틀과 말해서 창의적으로 생각하면서 왜 비판을 하던 한 번은가? 창의력의 기본 모태
2	18124703991370752	2014-01-01 00:07:22	08311nest1	LumSoftu_로스루리디게	ko	0	@bamduul_1 평가 창의적으로 이루어지지 않는다 될
3	181250026963755008	2014-01-01 00:21:34	Juan_bot	준인	ko	0	@heeking 개소리도 그렇고 창의력이라고 그래
4	181257584819159040	2014-01-01 01:08:54	죽은시인의시집	죽은시인의시집	ko	0	*호기심소년 안철수 창의력이라고 되다? 이런 디테일은 안 안철수와의 관련해서 시인은 의과대 입학
5	18126263707550657	2014-01-01 01:17:12	imgooxcom	디딤장병	ko	0	0 내 창의 비빌은 붉은보시였다
6	18126667334861824	2014-01-01 01:25:20	sejongsarang	임정신	ko	0	0 융합적 창의력(융합 창의력의 사람)을 발휘하라. 창의적인 생각을 이끌어내라
7	18126831373553024	2014-01-01 01:31:51	sejongsarang	임정신	ko	0	0 융합적 창의력(융합 창의력의 사람)을 발휘하라. 창의적인 생각을 이끌어내라
8	181272546136939872	2014-01-01 01:48:40	YoungSeKim	YoungSe Kim	ko	0	0 융합의 결과물의 가치는 서로 다른 분야의 만남 통해 극대화된다. 예상하기 쉽지 않은 분야와
9	181273339174312768	2014-01-01 01:51:49	YoungSeKim	YoungSe Kim	ko	0	0 심리학 유려하지 않고 남들이 하지 않은 분야에 도전하는 일은 마음을 편하게 그 후회하는 사
10	18127382289483776	2014-01-01 01:53:11	WILL_DARIN	DH	ko	0	0 @YoungSeKim 오늘 오전에 대한민국의 미래를 만나 뵙겠습니다! 역시 외환부 사람들입니다
11	181276470088151040	2014-01-01 02:04:13	pujdingcat	게으른 딤딤	ko	0	0 2014년은 좀더 창의적인 사람이 되고 싶어서(창의력)가득찬다
12	18127654188204417	2014-01-01 02:04:33	HYUbob	HYUbob	ko	0	0 (E) 창의력(창의성) : (아재개스)개, 갓말지못, 우동국, 김치, 밥(밥)
13	1812792584947328	2014-01-01 02:15:13	Bayrambooks	바이람(도서출판 프리라)	ko	0	0 [서평] '애플은 창의적 디지털 서비스 업체' (아이클라우드, 스티브 잡스가 디자인한 애플의 미래, ;
14	181280340377964546	2014-01-01 02:19:39	lan_Rumpus_Ct	남궁연	ko	0	0 @Chase_Alphado (주목으로 창의성을 꼭 돈드린다) (자강)
15	18128123209088832	2014-01-01 02:23:33	sejongsarang	임정신	ko	0	0 아이작의 '스티브 잡스' (융합 창의력의 사람)을 발휘하라 http://url.me/6awwf 창의력
16	181281817053112353	2014-01-01 02:20:45	jtbclove	JTBC	ko	0	0 중간 기술의 콘서트 티켓 판매 급증과 음원 불출, 콘서트 중계도까지 바꾸며 저속을 드러낸 #비
17	181283369610063872	2014-01-01 01:11:25	BonbonOn2	오즈	ko	0	0 번가람 대학원장의 알리스토크 데이비에서 발견-잊혀질수록, 오즈가 그리던 #오즈의 사랑
18	181283139498051584	2014-01-01 02:43:00	MoonwonJ_bot	최원웅 of actor JJ	ko	0	0 비즈니스 톨링 폰트와, 내용이 창의적인가 아닌지 알아차려, 서툰데, 그날 앞의 예기 할까? #seer
19	181283996689105920	2014-01-01 04:14:59	Seoul_140416	2주차 관람 14일 딱전 조	ko	0	0 Q. 창의적 그림을 그릴때 참고하는 소제는? 답변은 탄 베틀스가 산책도 후회를 다치면 가지
20	181283263277801472	2014-01-01 04:30:28	Karna_Jrbot	카르나	ko	0	0 ..영원처럼 복종하지, 나의 창의력을 담고, 그들을 불태우도록 하지, 내 안에 있는 것은 불꽃이
21	181283218064037888	2014-01-01 05:18:22	Huge_Ego	Ego	ko	0	0 귀양창의 지적인 노력처럼 찾지, 평타처럼, 정답은 '탄탄' 장난이나 pic.twitter.com/ABy2R12UK
22	181282632932264960	2014-01-01 05:22:31	quensagemeister	퀸스그대이스티	ko	0	0 더 창의적으로 살아가는 다섯가지 방법 http://buddyholly1.blog.me/13018277692
23	181283002866712576	2014-01-01 05:36:59	nyxly	jin	ko	0	0 구글클래스 기사를 보니 테드강의 단면 소문이 연상된다. 공극적인 라이프로그, http://nyxly.com/A
24	1812836876043321344	2014-01-01 06:04:18	JDB578	홍이생각통(洪)	ko	0	0 창의적인 사람들은 서로 다르고 하지만 한 가지 점에서 일치한다. 그것은 자신이 하는 일을 사랑한
25	181283740232068848	2014-01-01 06:06:33	HighForum	한국 고등학생 포럼	ko	0	0 한국 고등학생 포럼은 2세기 이래 상류층에 밀려 놓여왔던 창의적인 청소년 리더 양성을 목표로
26	1812844631096401912	2014-01-01 06:30:22	dragonj66	백운우	ko	0	0 [대이비] #비즈니스 디너쇼스튜디오, 문화창의학교 유치활동은 현재 중점 http://jme2do.kr/in?chc# 해설0
27	1812845145818206320	2014-01-01 06:37:09	Hyema_magic	강혜은	ko	0	0 @Hana_magic @kongma_magic 그런 만연이 아니라, 창의적인 인재는 그런 환경에 어울리지 않는
28	181284525274580993	2014-01-01 06:39:51	myongdae_LEE	myongdae, LEE	ko	0	0 한말한리 때는 60%의 공백은 여론과 39%의 창의력은 1%의 독재로 움직인다. ..권력자들의 수
29	181285283126744946	2014-01-01 07:04:31	ftfx5604	(인생...) 소우세취이	ko	0	0 @Blue_Lake_ 씨앗(안창의적
30	181285312728951936	2014-01-01 07:08:52	imgooxcom	디딤장병	ko	0	0 내 창의 비빌은 붉은보시였다
31	181285421759615840	2014-01-01 07:13:25	cyp_ttp_bot	사이피즈 사소한 팀뽕(홍)	ko	0	0 기상천외의 기상유적 시간은 보통 기상유적 시간의 반 정도 됩니다.
32	181285471399005184	2014-01-01 07:15:10	Midoriage	영미이다	ko	0	0 '자기실현 과정은 그 자체가 창의적인 과정이다. 자기의 지평을 모두 활용해서 중간기에 맞춘 삶
33	1812860487496005886	2014-01-01 07:38:07	r569	Nam Yeon Jeong	ko	0	0 @kehendardidid 해피뉴이얼 반박 ㅋㅋㅋ 창의력이다 ㅋㅋㅋ 오빠의 꿈 ㅋㅋ
34	181286102041418912	2014-01-01 07:43:33	daaP7	5준수장	ko	0	0 인간은 창의적인 동물이다. 내년 인간이 될 표기했나 보네. ..크게 답이 없다는 속속 뜻했...)
35	181286318018881744	2014-01-01 07:48:49	green_lee75	여기저기 아픈 시루 찰떡	ko	0	0 @Beam_sd 올라 난 작당 별로 안올아해서 잘안해서 막창의 심오함까지는 공감할수있어 ㅋㅋ
36	181286544097978368	2014-01-01 07:57:49	alwong98123	뽕따먹	ko	0	0 Somethine 최고!!! 어쨌든 그 창의적인것을 만두수있지. ...ㅋㅋㅋ 대박이고 노래도 너무 좋고 오빠

Figure 2. Collection of Semi-structured Data

3.2. Preprocessing Procedures for Data Collection

The historic data gathered by media performs the preprocessing process, and morpheme analysis and natural language process methods by decomposition in text mining are

required to analyze the semi-structured data on text documents. Unnecessary attribute factors of the data field are removed from the Korean sentences in the preprocessing procedures. For instance, in the case of collecting the historic data of Twitter, the attribute values of excel data are classified into Twitter numbers, date, screen name of the writer, name of the writer, language, number of re-tweets and contents, and accordingly, the preprocessing is carried out by extracting the content and date on which the writing was written by the writer. Twitter number & language, tweet time stamp and et cetera are removed, for the deduction of multiple tweets and promotional tweets or unclear tweets comprised of symbols other than letters have been suggested. Tweet text contents are extracted as a text file separately for tweets that do not fall under the category of the above mentioned case.

원본	대상	유형	Null 허용	크기	전체 자...	소수 ...
트윗번호	트윗번호	nvarchar	<input checked="" type="checkbox"/>	255		
날짜	날짜	nvarchar	<input checked="" type="checkbox"/>	255		
글쓴이 스크린 이름	글쓴이 스크린 이름	nvarchar	<input checked="" type="checkbox"/>	255		
글쓴이 이름	글쓴이 이름	nvarchar	<input checked="" type="checkbox"/>	255		
언어	언어	nvarchar	<input checked="" type="checkbox"/>	255		
리트윗갯수	리트윗갯수	float	<input checked="" type="checkbox"/>			
내용	내용	nvarchar	<input checked="" type="checkbox"/>	255		

Figure 3. Data Attribute Preprocessing

3.3. Preprocessing of Natural Language in Korean Texts

The process for the preprocessing of natural language is carried out on Twitter texts by using the Korean morphological analyzer. The sentences in Twitter texts are mixed with parts of speech, such as subjects, nouns, verbs, adjectives, adverbs, exclamations, etc., and such characteristics are not easy to analyze with the attributes of natural language. A three-phase process by natural language process is carried out to analyze this.

Table 1. The Three Phases of the Execution of Natural Language Preprocessing

Phases	Rules for processing the natural language
Step 1. Process non-Korean texts	Removal of characters excluding Korean, special symbols, unnecessary blanks, numbers, alphabets, etc.
Step 2. Filtering of Korean words by parts of speech	Assortment of parts of speeches that do not convert to the subject of analysis of unspecified subjects, meaningless adverbs, exclamations, postpositional particles
Step 3. Preprocessing of dictionary-based vocabularies	Vocabularies falling under the category of slangs, terminologies related to communications, alien words and et cetera are eliminated as meaningless non-dictionary vocabularies by comparing them to dictionary-based vocabularies.

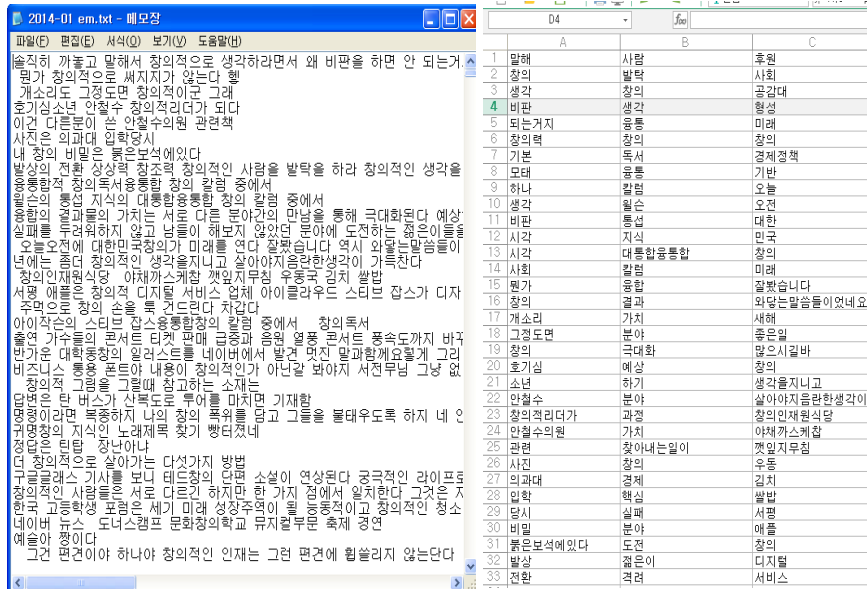


Figure 4. Result of the Execution of non-Korean Processing and Dictionary-based Preprocessing

4. Big Data Analysis: The Execution of Collective Examination of Twitter Time Series' Recurrent Entries

Frequent patterns are patterns that occur frequently in data sets. In other words, by analyzing the terminologies and sentences that appear most frequently in Twitter by having them applied to Twitter, most frequently used frequent keywords regarding creative education are extracted. In addition, by reflecting the fact that they are log-like records of short sentences, sentences and keywords in a long-range time period due to the attributes of Twitter, extraction of Twitter keywords and frequency time series analysis are carried out simultaneously for analysis.

4.1. Creation of Time Series Transactions for Twitter

Twitter is comprised of short sentences of unspecified individuals. In order to analyze the vocabularies in the message, it is necessary that vocabularies of the twit irrelevant to vocabularies existing in close proximity to twits made during the same period are classified. Therefore, there is a need to create the transactions by categorizing this into the same group and other group. In this thesis, time series transactions are constructed based on time, ID and et cetera pursuant to Twitter.

Table 2. Illustration of Time Series Twitter Transaction

인재.	교육.	교육혁명.	대통령.	생각.	사람.	창조.	미래.
방법.	혁신.	교육.	목표.	생각.	창조.	상상.	사람.
생각.	창조.	인재.	사람.	경제.	교육.	디자인.	체험.
사람.	생각.	창조.	방법.	가치.	디자인.	비밀.	경제.
사람.	디자인.	생각.	인재.	방법.	가지.	교육.	사업.
생각.	사람.	교육.	세상.	방문.	체험학습.	인간.	가지.
사람.	생각.	체험학습.	방문.	교육.	가지.	체험.	디자인.
사람.	가지.	체험학습.	방문.	생각.	일반.	행동.	디자인.
사람.	디자인.	생각.	교육.	체험학습.	가지.	방과후.	학교.
교육.	사람.	생각.	디자인.	아이디어.	창조.	경기.	체험학습
뉴스.	사람.	생각.	디자인.	교육.	체험학습.	창조.	방문.
사람.	가지.	생각.	체험.	습관.	디자인.	광고.	교육.
최창.	광고.	생각.	모음.	디자인.	교육.	체험.	사람.
광고.	모음.	생각.	교육.	사람.	방문.	체험학습.	디자인.
교육.	이재.	공공.	계획.	시장.	조리.	청렴.	산후.
광고.	모음.	교육.	생각.	디자인.	융합.	사람.	가지.
사람.	가지.	생각.	광고.	모음.	디자인.	교육.	문창.
.....

4.2. Analysis of Support and Confidence of Keywords in Twitter

The number of keywords of twits collected from the historic data is enormous, and in order to extract meaningful keywords from them, the three execution phases of frequency analysis, support analysis and confidence analysis must be carried out.

(i) Frequency Analysis

Frequency analysis, by priority, filters those vocabularies that are twitted in lesser amount. In other words, vocabularies that are not mentioned frequently in the whole data are eliminated, and on the contrary, vocabularies that are mentioned frequently become primary keywords for finding associated words and analyzing their meanings.

(ii) Support Analysis

Support analysis, similar to frequency analysis, are used to measure the frequency rate of the vocabularies from the entire twits. In this thesis, the support levels of twits entered by transactions are measured by each twit, while excluding the keywords that are repeated, duplicated and misspelled within a single twit. In addition, it measures the relative support levels between the keywords with support levels by twit as a basis. Only the whole data number ratio was considered in the previously existing data mining method, but since it has the problem of not being able to estimate the whole data number if the target is set to be Twitter, relative support levels of partial twits must be considered. In other words, twits that are related to ‘creative education’ were set as the target in this thesis, support levels were measured on the basis of twits that have appeared in connection with relevant keywords, and with that measurement as foundation, the validity of keywords are assessed.

(iii) Confidence analysis

The reliability analysis of twits is carried out through extracting the vocabularies directly relevant with each other and by the measurement of figuring out the relations of the vocabularies.

4.3. Analysis of Valid Twitter Keywords

The analysis of valid keywords is a process for extracting relevant, meaningless and meaningful keywords only amongst vocabularies that haven't gone through the preprocessing process. Only the high ranking keywords are refined after frequency, support and reliability are reviewed. Here, the criteria for evaluating the validity of keywords are as the following.

$$\text{keyword effectiveness} = \frac{F_{(\text{frequency})}(\text{Keyword})}{\text{Number of transactions} * \text{Maximum}^{\text{port}} \text{ of Keywords}} * \text{Maximum} | C_{(\text{confidence})}(\text{with Other Keywords})$$

The refinement result of vocabularies used simultaneously with vocabularies relevant to creative education are as the following.

인재	창조	문제	에너지
227	214	71	70
사상	경제	가지	연대
206	148	68	68
아이디어	체험	이벤트	대박
134	132	68	67
방법	디자인	하계	학기
131	130	67	67
붉은보석에있다	비밀	걱정	행일
124	124	65	63
미래	혁신	거대	비홍
122	117	62	61
목표	자신	용복	국부
115	113	61	60
대통령	학교	시간	아이
106	104	60	60
교육혁명	필요	추구	공장
98	96	60	59
체험학습	상상	기업	달나라
90	88	59	59
청소년	우리	문샷	사랑
86	83	59	59
몰이	발상	소수	실제
82	82	59	59
시각	자기	우주	가변
81	79	59	59
가능	기상	가슴	고체
74	73	58	58

Figure 5. Result of the Refinement of Twitter Keywords

4.4. Refinement of Result Keywords

The result keywords are shown by using visualization tools after selecting keywords that are of higher validity in consideration of the semanticity to the related words.

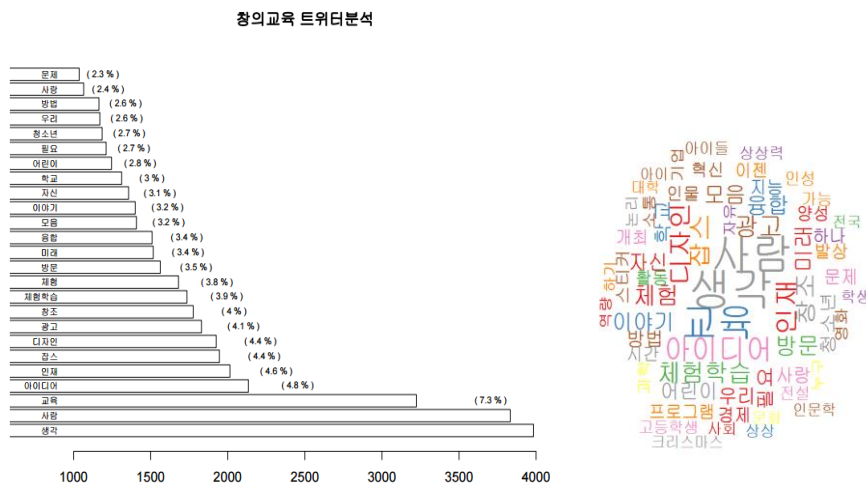


Figure 6. Analysis of Creative Education Twitter

5. Performance Evaluation

In this thesis, meaningful keywords that are suggested on Twitter have been extracted and their frequencies have been measured by using the two keywords, ‘creative’ and ‘creative education’, and accordingly, how performance and issues of creative education were reflected was analyzed. The analyzation of the measurement on dynamization, issues and performance factors of creative education with Korean Twitter as target, which was carried out with semi structured data written by internet users, has shown that, for the case of keywords related to ‘creative’, the influences of the word ‘education’ has increased to a great extent as the significance of creative education policies has been highlighted with the coming of the new administration to power, and the word education, even in year 2014, has continued to show a steady uptrend. However, it has been analyzed that the word ‘software’ has been highlighted in year 2015 by the spread of software education, which was the objective of the creative human resource education.

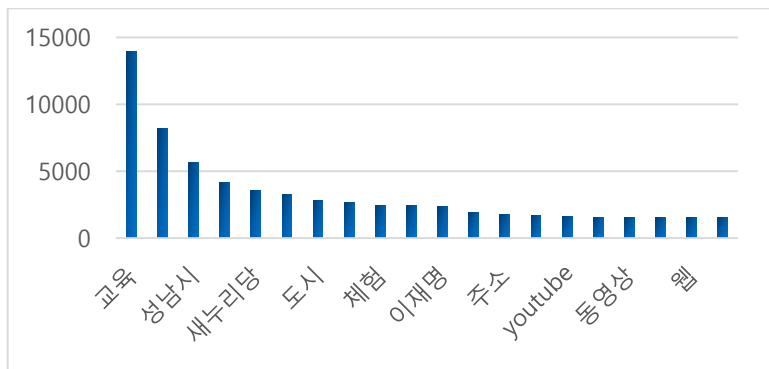


Figure 7. 2013 Creative Keyword Top 20

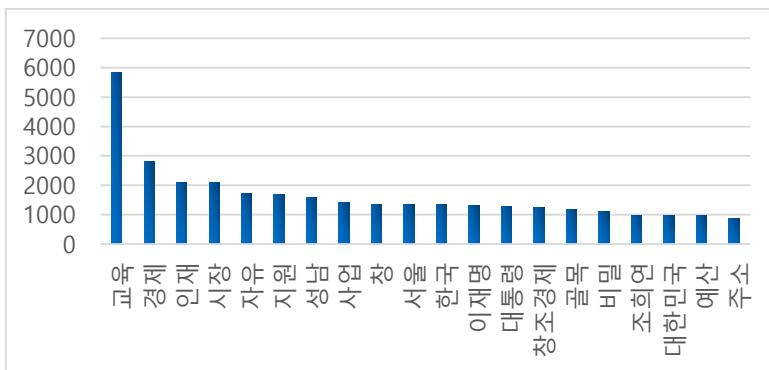


Figure 8. 2014 Creative Keyword Top 20

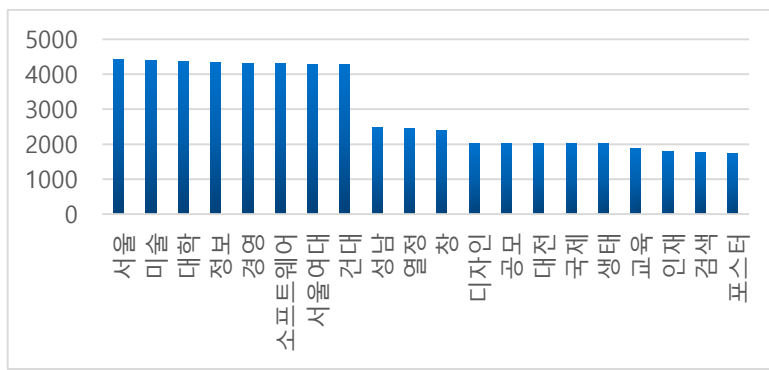


Figure 9. 2015 Creative Keyword Top 20

Following is the result of the analysis of keywords of 'creative education'. With regard to the time period, tweets that were made between year 2014 and 2015 were collected. The ground for analyzing the keywords of 'creative education' from year 2014 is in the fact that the phrase 'creative education' has not been widely established in South Korea in year 2013 and the deductions can be made from Twitter were minimal, and thus it has been considered as a factor likely to lower the validity of the analysis.

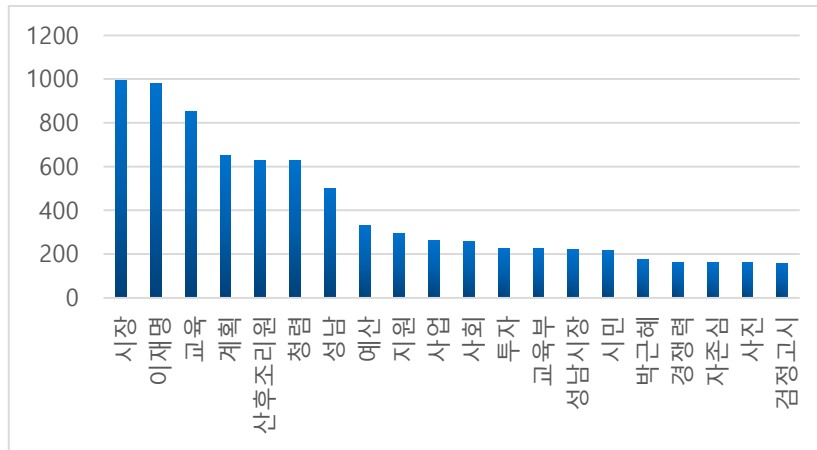


Figure 10. 2014 Creative Education Keyword Top 20

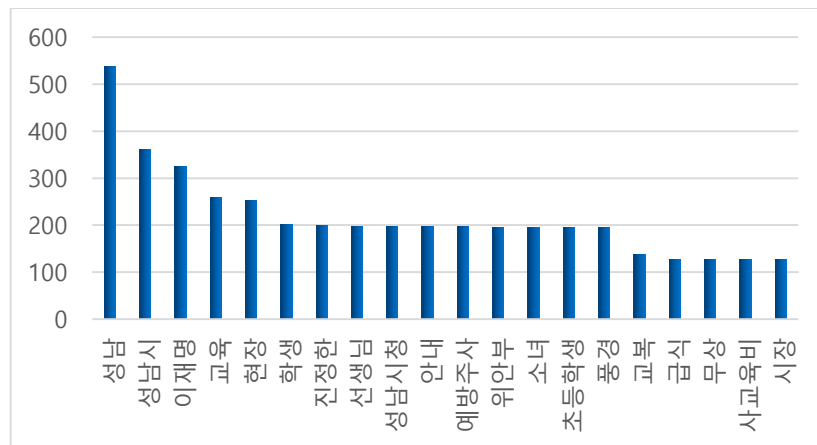


Figure 11. 2015 Creative Education Keyword Top 20

As a result of the analysis made, it has been confirmed that concerns regarding private education expenses and budget on national policies and elementary school students were highlighted as a social issue in South Korea, and, regardless of the fact that notable achievements regarding 'creative education' have not yet been made, it has been analyzed that various areas of education, including but not limited to arts, music and physical education, are conducting the creative education.

6. Conclusions

This thesis, on the basis of semi-structured data of Twitter comprised of 140 words, had collected the tweet data ranging from year 2013 to 2015 and drawn a conclusion, through big data analyzation, on the measurement and performance factors on the subject of to what degree had creative education permeated the Korean educational system. Despite the fact that the so-called creative education is not yet prevalent in the Korean educational system in a broad perspective, it has been confirmed that it is exercising a huge influence in the

Korean society in general via fragments of their national policies, software education and et cetera. However, there lies a need and challenge of pursuing further extensive studies through various statistical analysis other than the frequency analysis used in this thesis, such as data mining, text mining, opinion mining and et cetera.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education, Science Technology(NRF-2014R1A1A1008524).

References

- [1] Y. Sun and K. Jia, "Research of word sense disambiguation based on mining association rules", Third International Symposium on Intelligent Information Technology Application workshops, NanChang, China, (2009), pp. 86-88.
- [2] M.H. Dunham, "Data Mining: Introductory and Advanced Topics", Pearson Education Inc., (2003).
- [3] C. Tang, and C. Liu, "Method of Chinese grammar rules automatically access based on association rules", Proceedings of the. Computer Science and Computational Technology, vol. 1, (2008), pp. 265-268.
- [4] Y. Xu, Y. Li & G. Shaw, "Reliable representations for association rules", Data & Knowledge Engineering, vol. 70, no. 6, (2011), pp. 555-575.
- [5] Y. S. Chae and K. S. Choi, "The KIBS project: KAIST corpus and part-of-speech and syntactic tagset", Proceedings of workshop on multi-lingual information processing and Asian language processing (MAL'99), serial. 1, (1999), pp. 19-22.
- [6] D.W. Choi and Y.J. Hyun, "Transitive association rule discovery by considering strategic importance", 10th IEEE International Conference on Computer and Information Technology, (2010), pp. 1654-1659.
- [7] D. Jurafsky and J. H. Martin, Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, Prentice-Hall, Upper Saddle River, NJ, (2000).
- [8] I.A. Khan and J.T. Choi. "An application of educational data mining (EDM) technique for scholarship prediction", International Journal of Software Engineering and its Applications, vol.8, no.12, (2014), pp. 31-42.
- [9] B. Pang, L. Lee and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques", Proceedings of the ACL-02 Conference on Empirical methods in Natural Language Processing, vol.10, (2002), pp.79-86
- [10] H. Ming, N. Wenying, and L. Xu, "An improved decision tree classification algorithm based on ID3 and the application in score analysis", Chinese Control and Decision Conference (CCDC), (2009), pp. 1876-1879.
- [11] M. Kantardzic, "Data Mining: Concepts, Models, Methods and Algorithms", Wiley-IEEE Press. International Journal of Data Mining & Knowledge Management Process (IJDMP), vol.3, no.5, (2013).

Authors



Kil Hong Joo, He received the M.S. and Ph.D. degree in Computer Science from Yonsei University, Seoul, Korea, in 2000 and 2004. He is currently a professor of Department of Computer Education at Gyeongin National University of Education, Korea. His current interests include mining data streams, data analysis and smart learning.



Seo-Ji Hoon, He received his Bachelor degree in 2008 at Seoul National University of Science and Technology from department of Safety Engineering. He finished his MS and Ph.D. at Incheon National University from Department of Computer Science and Engineering in 2010 and 2015 respectively. His research interest includes Data Mining, Database Management and Sensor Networking.



Nam Hun Park, He received the B.S., M.S. and Ph.D. degree in Computer Science from Yonsei University, Seoul, Korea, in 2000, 2002 and 2007. He was a post-Ph.D. at the Department of Computer Science, Worcester Polytech Institute, Worcester, MA. He is currently a professor of Department of Computer Science at Anyang University, Korea. His current interests include mining data streams.

