ARTIFICIAL INTELLIGENCE IN 1,000 KCI ARTICLES: A BIG DATA ANALYSIS

Namkil Kang
Far East University, South Korea

This paper aims to analyze the abstracts of 1,000 KCI (Korea citation index) articles published from 2015 to 2023 regarding artificial intelligence. A point to note is that the eight noun-sentence was the most widely used in 1,000 abstracts. A major point of this paper is that in the word cloud of 1,000 abstracts, the noun *intelligence* is deemed to be the most centric, followed by the word *AI*, the noun *technology*, the noun *study*, and the noun *datum*, in descending order. A further point to note is that topic 17 was the most frequently used in 1,000 abstracts, followed by topic 9, topic 7, and topic 14, in that order. With respect to the frequency of the so-called nouns, it is interesting to point out that the noun *intelligence* was the most widely used in 1,000 abstracts, followed by the word *AI*, the noun *study*, the noun *technology*, and the noun *result*, in descending order. When it comes to the map of closeness centrality, the word *AI* is the highest in the closeness centrality. Simply put, it showed up in the center of the map, thereby implying that it is the most centric and central.
Keywords: KCI articles, topic, token, keyword, frequency, big data, abstract

1. Introduction

The main goal of this paper is to provide an in-depth analysis of the abstracts of 1,000 KCI (Korea citation index) articles published from 2015 to 2023 regarding artificial intelligence. We collected and analyzed them in terms of the biblio data collector and the software package NetMiner. First, we aim at probing into the frequency of the so-called nouns that occurred in 1,000 abstracts. More specifically, we inquire into their frequency results, their proportion, and their cumulative proportion. Second, we aim to provide the word cloud of 1,000 abstracts in which representative keywords show up and they are all different in fonts. Put differently, more pivotal words turn up in bigger words. Third, we attempt to look into 17 topics that represent 1,000 abstracts and their keywords. Additionally, we aim to examine which topics showed up more frequently in 1,000 abstracts. Also, we attempt to contemplate the map of 17 topics and their keywords in which central keywords are linked to 17 topics that represent 1,000 abstracts. Fourth, we aim at going over the use of the key words that turned up in the abstracts of 1,000 KCI articles. Finally, attention is paid to the map of closeness centrality that refers to the distance between words. The more the distance between words is close, the more they are deemed to be centric and pivotal in the map. The terminology centrality refers to the so-called influence, prestige, importance, and prominence. Simply put, if the distance between words is close, then they are more influential and important in the map.

2. Results

2.1. Word Frequency

This section centers on providing the so-called frequency of nouns that occurred in 1,000 abstracts including their proportion and cumulative proportion:

Table 1 Noun frequency

Value	Frequency	Proportion	Cumulative Proportion
1.0	4,359	0.532	0.532
2.0	1,123	0.137	0.669
3.0	562	0.069	0.738
4.0	378	0.046	0.784
5.0	250	0.031	0.815
6.0	160	0.02	0.834
7.0	138	0.017	0.851
8.0	98	0.012	0.863
9.0	108	0.013	0.876
10.0	65	0.008	0.884
11.0	60	0.007	0.892
12.0	54	0.007	0.898
13.0	57	0.007	0.905
14.0	42	0.005	0.91
15.0	32	0.004	0.914
16.0	23	0.003	0.917
17.0	29	0.004	0.921
18.0	28	0.003	0.924

19.0	31	0.004	0.928
20.0	21	0.003	0.93
21.0	18	0.002	0.932
22.0	19	0.002	0.935
23.0	18	0.002	0.937
24.0	15	0.002	0.939
25.0	16	0.002	0.941
26.0	14	0.002	0.942
27.0	18	0.002	0.945
28.0	12	0.001	0.946
29.0	15	0.002	0.948
30.0	21	0.003	0.951
31.0	17	0.002	0.953
32.0	12	0.001	0.954
33.0	9	0.001	0.955
34.0	14	0.002	0.957
35.0	5	0.001	0.958
36.0	14	0.002	0.959
37.0	8	0.001	0.96
38.0	5	0.001	0.961
39.0	8	0.001	0.962
40.0	4	0	0.962

41.0	8	0.001	0.963
42.0	5	0.001	0.964
43.0	5	0.001	0.964
44.0	5	0.001	0.965
45.0	4	0	0.966
46.0	10	0.001	0.967
47.0	10	0.001	0.968
48.0	3	0	0.968
49.0	6	0.001	0.969
50.0	4	0	0.97
51.0	5	0.001	0.97
52.0	5	0.001	0.971
53.0	5	0.001	0.971
54.0	7	0.001	0.972
55.0	3	0	0.973
56.0	4	0	0.973
57.0	3	0	0.974
58.0	3	0	0.974
59.0	3	0	0.974
60.0	4	0	0.975
61.0	3	0	0.975
62.0	3	0	0.975

63.0	5	0.001	0.976
65.0	1	0	0.976
66.0	5	0.001	0.977
67.0	5	0.001	0.977
69.0	2	0	0.978
70.0	5	0.001	0.978
71.0	4	0	0.979
72.0	3	0	0.979
73.0	3	0	0.979
74.0	2	0	0.98
75.0	1	0	0.98
76.0	2	0	0.98
77.0	3	0	0.98
78.0	1	0	0.981
80.0	2	0	0.981
81.0	3	0	0.981
82.0	2	0	0.981
83.0	3	0	0.982
84.0	2	0	0.982
85.0	1	0	0.982
86.0	3	0	0.983
87.0	4	0	0.983

88.0	2	0	0.983
89.0	2	0	0.984
90.0	1	0	0.984
92.0	4	0	0.984
93.0	3	0	0.984
94.0	1	0	0.985
96.0	1	0	0.985
97.0	4	0	0.985
98.0	1	0	0.985
99.0	1	0	0.985
100.0	1	0	0.986
101.0	1	0	0.986
102.0	2	0	0.986
105.0	2	0	0.986
106.0	3	0	0.987
107.0	4	0	0.987
108.0	1	0	0.987
109.0	2	0	0.987
110.0	3	0	0.988
112.0	1	0	0.988
113.0	1	0	0.988
114.0	1	0	0.988

116.0	1	0	0.988
117.0	1	0	0.988
118.0	2	0	0.989
119.0	3	0	0.989
120.0	1	0	0.989
121.0	3	0	0.989
123.0	1	0	0.99
125.0	2	0	0.99
126.0	1	0	0.99
127.0	2	0	0.99
128.0	1	0	0.99
129.0	1	0	0.99
130.0	1	0	0.991
131.0	2	0	0.991
134.0	1	0	0.991
139.0	1	0	0.991
141.0	1	0	0.991
143.0	1	0	0.991
145.0	1	0	0.991
146.0	1	0	0.992
150.0	1	0	0.992
152.0	3	0	0.992

155.0	1	0	0.992
156.0	1	0	0.992
157.0	2	0	0.993
164.0	1	0	0.993
165.0	2	0	0.993
166.0	1	0	0.993
168.0	2	0	0.993
169.0	1	0	0.993
172.0	1	0	0.994
173.0	1	0	0.994
175.0	1	0	0.994
177.0	1	0	0.994
180.0	1	0	0.994
182.0	1	0	0.994
183.0	1	0	0.994
188.0	1	0	0.994
189.0	1	0	0.995
190.0	1	0	0.995
192.0	1	0	0.995
197.0	1	0	0.995
200.0	1	0	0.995
206.0	1	0	0.995

207.0	1	0	0.995
208.0	1	0	0.995
209.0	2	0	0.996
215.0	2	0	0.996
222.0	1	0	0.996
223.0	1	0	0.996
224.0	1	0	0.996
232.0	1	0	0.996
260.0	1	0	0.996
267.0	1	0	0.997
269.0	1	0	0.997
276.0	1	0	0.997
277.0	1	0	0.997
313.0	1	0	0.997
315.0	1	0	0.997
322.0	1	0	0.997
327.0	1	0	0.997
328.0	1	0	0.998
337.0	1	0	0.998
354.0	1	0	0.998
356.0	1	0	0.998
372.0	1	0	0.998

386.0	1	0	0.998
408.0	1	0	0.998
426.0	1	0	0.998
453.0	1	0	0.999
454.0	1	0	0.999
477.0	1	0	0.999
488.0	1	0	0.999
520.0	1	0	0.999
538.0	1	0	0.999
550.0	1	0	0.999
740.0	1	0	0.999
864.0	1	0	1
908.0	1	0	1
1068.0	1	0	1
2063.0	1	0	1
3037.0	1	0	1
Total	8,189	1	

It is quite interesting to mention that one word showed up 4,359 times in 1,000 abstracts. Put differently, it obtained 4,359 tokens and the highest proportion (0.532). That is to say, it accounts for 53.2%. It is also very interesting that two words turned up 1,123 times in 1,000 abstracts and that their proportion is 0.137 (13.7%: the second highest). Quite interestingly, three words occurred 562 times in 1,000 abstracts. It is the third highest and accounts for 6.9% among 8,189 nouns. Particularly noteworthy is that four words appeared 378 times (the fourth highest). Their proportion is 0.046 (4.6% among 8,189 nouns) and their cumulative proportion is 0.784. It must be stressed that five words showed up 250 times in 1,000 abstracts (the fifth highest). Their proportion is 0.031 (3.1%) and their cumulative proportion is 0.815. We thus

conclude that one word was the most frequently used (4,359 tokens) and obtained the highest proportion (53.2%) among 8,189 nouns.

Now attention is paid to word length:

Table 2 Word length

Value	Frequency	Proportion	Cumulative Proportion
2.0	158	0.019	0.019
3.0	410	0.05	0.069
4.0	630	0.077	0.146
5.0	704	0.086	0.232
6.0	781	0.095	0.328
7.0	809	0.099	0.426
8.0	823	0.101	0.527
9.0	782	0.095	0.622
10.0	702	0.086	0.708
11.0	542	0.066	0.774
12.0	392	0.048	0.822
13.0	320	0.039	0.861
14.0	194	0.024	0.885
15.0	162	0.02	0.905
16.0	110	0.013	0.918
17.0	76	0.009	0.927
18.0	57	0.007	0.934
19.0	48	0.006	0.94

20.0	34	0.004	0.944
21.0	26	0.003	0.948
22.0	35	0.004	0.952
23.0	32	0.004	0.956
24.0	21	0.003	0.958
25.0	12	0.001	0.96
26.0	12	0.001	0.961
27.0	20	0.002	0.964
28.0	18	0.002	0.966
29.0	18	0.002	0.968
30.0	25	0.003	0.971
31.0	22	0.003	0.974
32.0	21	0.003	0.976
33.0	14	0.002	0.978
34.0	19	0.002	0.98
35.0	12	0.001	0.982
36.0	17	0.002	0.984
37.0	11	0.001	0.985
38.0	7	0.001	0.986
39.0	11	0.001	0.988
40.0	9	0.001	0.989
41.0	8	0.001	0.99

42.0	10	0.001	0.991
43.0	9	0.001	0.992
44.0	3	0	0.992
45.0	11	0.001	0.994
46.0	2	0	0.994
47.0	3	0	0.994
48.0	7	0.001	0.995
49.0	1	0	0.995
50.0	2	0	0.995
51.0	4	0	0.996
52.0	2	0	0.996
53.0	8	0.001	0.997
54.0	2	0	0.997
55.0	2	0	0.998
56.0	5	0.001	0.998
57.0	1	0	0.998
60.0	1	0	0.999
61.0	1	0	0.999
63.0	1	0	0.999
64.0	1	0	0.999
65.0	1	0	0.999
67.0	2	0	0.999

68.0	1	0	0.999
69.0	1	0	1
71.0	1	0	1
73.0	1	0	1
84.0	1	0	1
89.0	1	0	1
Total	8,189	1	

It is particularly noteworthy that the eight noun-sentence showed up 823 times in 1,000 abstracts (the highest). Its proportion is 0.101 (10.1%) among 8,189 nouns and its cumulative proportion is 0.527. It is also significant that the seven noun-sentence turned up 809 times in 1,000 abstracts (the second highest). Interestingly, its proportion is 0.099 (9.9%) among 8,189 nouns and its cumulative proportion is 0.426. It is appropriate to mention that the nine noun-sentence occurred 782 times (the third highest). It accounts for 9.5% and its cumulative proportion is 0.622. Finally, it must be noted that the six noun-sentence turned up 781 times in 1,000 abstracts (the fourth highest). Its proportion is 0.095, which in turn indicate 9.5%. It therefore seems reasonable to mention that the eight noun-sentence was the most widely used, followed by the seven noun-sentence, the nine noun-sentence, and the six noun-sentence, in that order.

Now let us turn our attention to Table 3:

Table 3 Common nouns and proper nouns

Value	Frequency	Proportion	Cumulative Proportion
Common Noun	4,732	0.578	0.578
Proper Noun	3,457	0.422	1
Total	8,189	1	

It is worth noticing that common nouns showed up 4,732 times among 8,189 nouns. Put differently, they account for 57.8%. It is also worth mentioning that proper nouns turned up 3,457 times among 8,189 nouns. More specifically, they account for 42.2%. We thus conclude that the frequency result of common nouns is by far higher than that of proper nouns in 1,000 abstracts.

2.2. The word cloud of 1,000 abstracts

The current section focuses on the word cloud of the abstracts of 1,000 KCI articles. Figure 1 shows the word cloud of 1,000 abstracts:

Figure 1 The word cloud of 1,000 abstracts

It is worthwhile pointing out that the noun *intelligence* is deemed to be the most centric and central since it is the biggest. We take that as suggesting that the word *intelligence* was the most occurred one in 1,000 abstracts. It is important to mention that the word *AI* is considered to be one of the most centric and central words. We take that as implying that its frequency is the second highest among the words of 1,000 abstracts. It should be pointed out that the noun *technology* counts as one of the most centric words. Quite interestingly, the noun *study* is assumed to be one of key words since it is the fourth biggest. It must be emphasized, on the other hand, that the noun *datum* is regarded as one of pivotal words since it is the fifth biggest. It therefore seems reasonable to mention that the noun *intelligence* is deemed to be the most centric, followed by the word *AI*, the noun *technology*, the noun *study*, and the noun *datum*, in descending order.

2.3. 17 topics and their keywords

The current section is devoted to inquiring into 17 topics and their keywords. Table 4 shows 17 topics and keywords that occurred in 1,000 abstracts:

Table 4 17 topics and their keywords

	1st Keyword	2nd Keyword	3rd Keyword	4th Keyword	5th Keyword
Topic-1	AI	computer	Artificial Intelligence	news	article
Topic-2	system	network	technique	control	power

Topic-3	Intelligence	paper	Artificial	approach	theory
Topic-4	datum	model	analysis	prediction	result
Topic-5	process	method	study	analysis	software
Topic-6	intelligence	concept	being	era	human
Topic-7	technology	field	development	intelligence	research
Topic-8	learning	machine	model	design	algorithm
Topic-9	disease	diagnosis	research	patient	review
Topic-10	intelligence	law	liability	human	responsibility
Topic-11	information	service	technology	study	industry
Topic-12	image	performance	result	model	algorithm
Topic-13	intelligence	work	law	art	issue
Topic-14	intelligence	ethic	problem	risk	value
Topic-15	study	analysis	research	result	factor
Topic-16	robot	development	person	right	society
Topic-17	education	study	student	intelligence	program

It is worth observing that the keywords *intelligence*, *paper*, *Artificial*, *approach*, and *theory* consist of topic 3. This in turn implies that they are centric since they are much used in 1,000 abstracts. It must be said, on the other hand, that topic 7 includes the keywords *technology*, *field*, *development*, *intelligence*, and *research*. This in turn indicates that they count as frequently used in 1,000 abstracts. More importantly, the keywords *image*, *performance*, *result*, *model*, and *algorithm* constitute topic 12. The noun *algorithm* is the fifth keyword, which in turn means that it is the fifth most centric in topic 12. What is interesting is that topic 14 includes the keywords *intelligence*, *ethic*, *problem*, *risk*, and *value*. What these keywords suggest is that artificial intelligence makes our life more convenient, but it has a side effect such as ethic problems or unemployment.

Now attention is paid to Table 5. Table 5 shows how widely 17 topics were used:

Table 5 Use of 17 topics

	Use of 17 topics
Topic-1	40
Topic-2	62
Topic-3	48
Topic-4	49
Topic-5	35
Topic-6	68
Topic-7	78
Topic-8	39
Topic-9	92
Topic-10	58
Topic-11	55
Topic-12	59
Topic-13	62
Topic-14	73
Topic-15	41
Topic-16	47
Topic-17	94

It is worthwhile noting that topic 17 showed up 94 times in 1,000 abstracts (the highest). As observed earlier, the keywords *education*, *study*, *student*, *intelligence*, and *program* are made up of topic 17. Quite interestingly, topic 9 turned up 92 times, which is the second highest. As exemplified in Table 4, topic 9 includes the keywords *disease*, *diagnosis*, *research*, *patient*, and *review*. Simply put, these keywords are about machine learning and it detects our disease correctly and easily. It should be noted that topic 7 occurred 78 times in 1,000 abstracts. As illustrated in Table 4, the keywords *technology*, *field*, *development*, *intelligence*, and *research*

constitute topic 7. Finally, it is worth noting that topic 11 appeared 73 times in 1,000 abstracts. As indicated in Table 4, topic 11 includes the keywords *intelligence*, *ethic*, *problem*, *risk*, and *value*. As mentioned earlier, these keywords tell us about the side effect of artificial intelligence. We thus conclude that topic 17 was the most frequently used in 1,000 abstracts, followed by topic 9, topic 7, and topic 14, in that order.

Now take a look at Figure 2:

Figure 2 Map of 17 topics
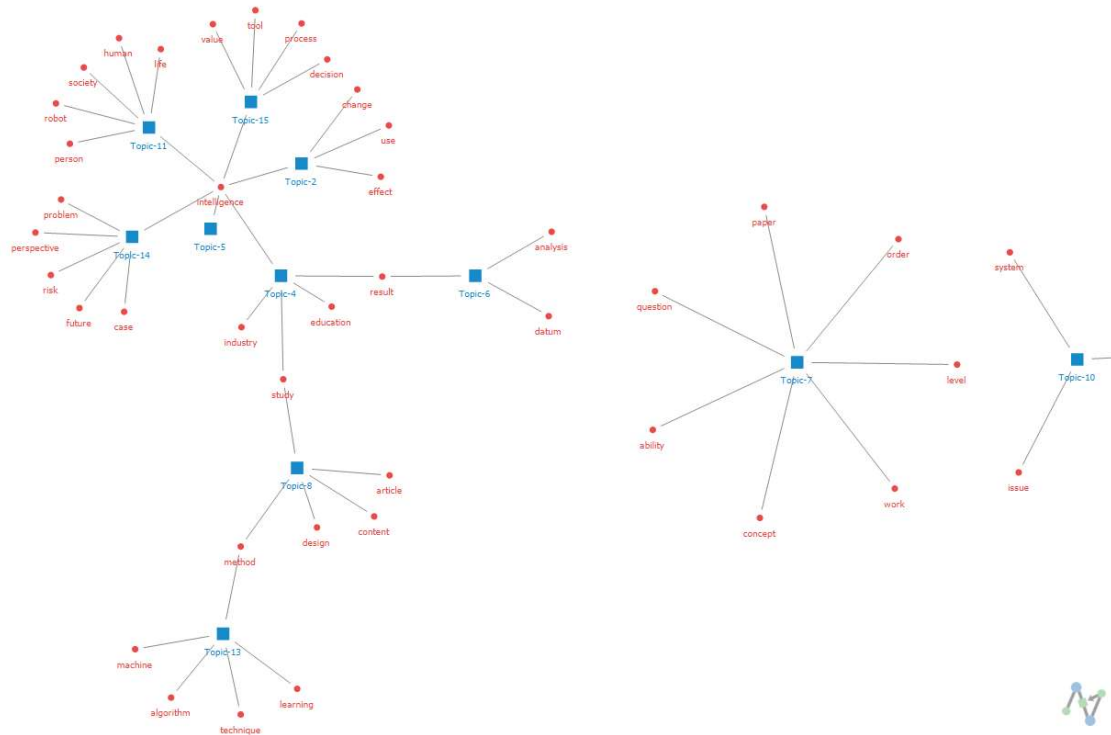
It is interesting to point out that topic 2, topic 5, topic 4, topic 11, topic 15, and topic 14 have the word *intelligence* in common. We take that as meaning that the keyword *intelligence* is the most centric and pivotal in 1,000 abstracts. We note that the keywords *life*, *human*, *society*, *robot*, and *person* are linked to topic 11. This seems to suggest that artificial intelligence makes our life easier and more convenient. Notice that the keywords *problem*, *perspective*, *risk*, *future*, and *case* are all linked to topic 14. More interestingly, these keywords seem to suggest that artificial intelligence gives us something convenient and valuable, but it takes away humanity. We thus conclude that six topics have the keyword *intelligence* in common, which is deemed to be the most centric.

2.4. Frequency of nouns

In what follows, we aim at examining the frequency of central words that occurred in 1,000 abstracts. Table 6 shows the frequency results of 33 central words:

Table 6 Frequency of central words

| Number	Word	Frequency
1	intelligence	805
2	AI	495
3	study	456
4	technology	434
5	result	349
6	datum	337
7	paper	328
8	system	318
9	development	314
10	research	295
11	method	257
12	analysis	251
13	Artificial	236
14	field	234
15	learning	218
16	problem	202
17	model	196
18	application	193
19	process	191
20	machine	184
21	information	182
22	algorithm	182
23	use	180
24	issue	167
25	time	164
26	future	162
27	addition	157
28	order	148
29	area	145
30	Artificial Intelligence	142
31	purpose	140
32	case	137
33	education	136

It is noteworthy that the noun *intelligence* has the highest frequency (805 tokens) in 1,000 abstracts. Simply put, the noun showed up 805 times in 1,000 abstracts. This in turn indicates that the authors of 1,000 articles used the noun *intelligence* the most frequently in 1,000 abstracts. What is interesting is that the word *intelligence* is followed by the word *AI*. To be more specific, the latter turned up 495 times in 1,000 abstracts. This is the second highest among the key words listed in descending order. It is important to note that the word *AI* is followed by the noun *study*. More specifically, the latter occurred 456 times in 1,000 abstracts.

It is interesting to point out that the noun *technology* appeared 434 times in 1,000 abstracts. This amounts to saying that artificial intelligence refers to cutting-edge technology. It must be noted that the noun *result* showed up 349 times in 1,000 abstracts. This in turn implies that this noun is one of the most frequently used words in 1,000 abstracts. It is vital that the noun *intelligence* was the most widely used in 1,000 abstracts, followed by the word *AI*, the noun *study*, the noun *technology*, and the noun *result*, in descending order. It is also worth mentioning that the noun *machine* occurred 184 times in 1,000 abstracts. In the term artificial intelligence, machine learning is one of the most pivotal features, hence obtaining the high frequency. We thus conclude that the noun *intelligence* was the most commonly used in 1,000 abstracts.

2.5 Closeness centrality

In the following, we aim to provide the so-called closeness centrality and its map. The term closeness centrality indicates that the more the distance between words is close, the more the words are important and centric. Figure 3 is the map of closeness centrality representing 1,000 abstracts. We eliminated many unnecessary nodes for the reason of space and included core words:

Figure 3 Closeness centrality
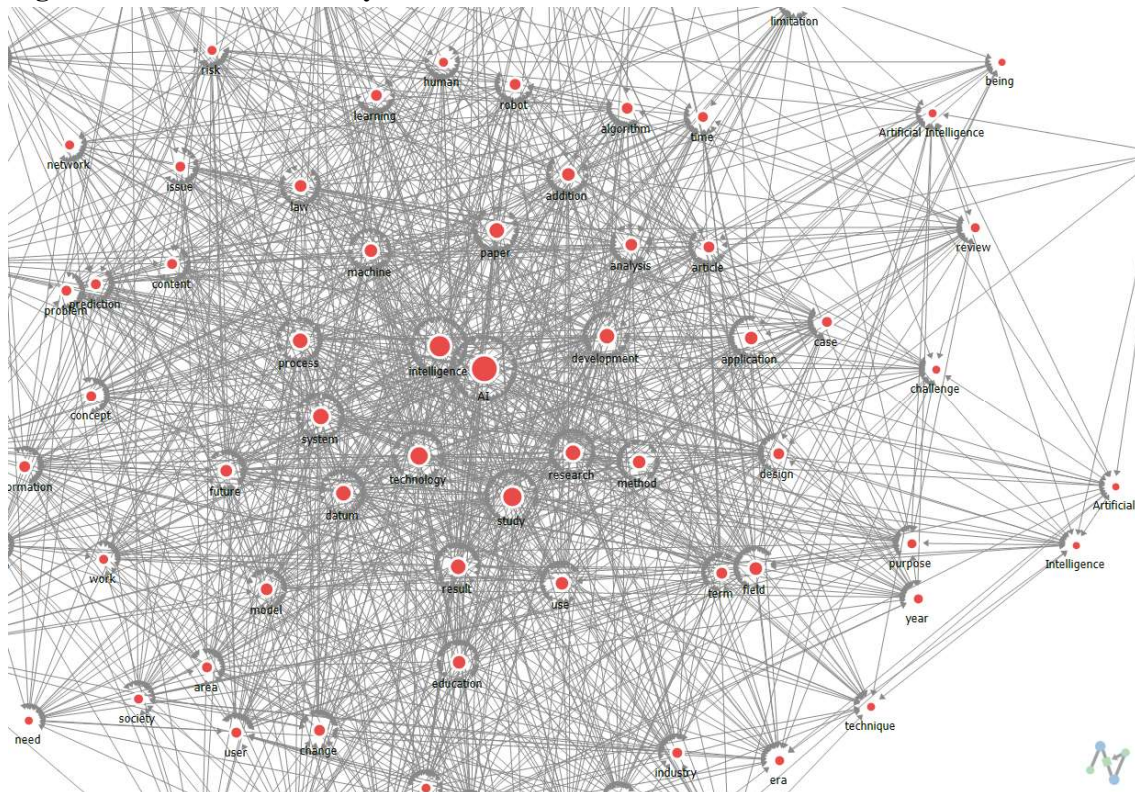
It is worthwhile to consider the word *AI*. When it comes to *AI*, it is the highest in the closeness centrality. More specifically, it showed up in the center of the map, thereby implying that it is the most centric and pivotal. In Figure 3, the word *AI* is the nearest to the other words. It is interesting to mention that the word *AI* is followed by the noun *intelligence*. That is to say, it is the second highest in the closeness centrality. It turned up around the center and thus it is

one of the most centric words, hence a core word. It is significant to note that the noun *intelligence* is followed by the noun *study*. The latter is the third highest in the closeness centrality. It also showed up in the center, namely near the word *AI*, hence counting as kind of centric. It is important to mention that the noun *technology* turned up near the word *study*. This in turn suggests that the word *technology* is kind of core in the closeness centrality. It seems thus appropriate to mention that the word *AI* is the highest in the closeness centrality, followed by the noun *intelligence*, the noun *study*, and the noun *technology*, in that order. For the maps of big data, see Kang (2023a, 2023b, 2023c, 2023d, 2023e, 2023f).

3. Conclusion

To sum up, we have analyzed the abstracts of 1,000 KCI (Korea citation index) articles published from 2015 to 2023 regarding artificial intelligence. In section 2.1, we have shown that the eight noun-sentence was the most widely used in 1,000 abstracts. In section 2.2, we have argued that in the word cloud of 1,000 abstracts, the noun *intelligence* is deemed to be the most centric, followed by the word *AI*, the noun *technology*, the noun *study*, and the noun *datum*, in descending order. In section 2.3, we have further argued that topic 17 was the most frequently used in 1,000 abstracts, followed by topic 9, topic 7, and topic 14, in that order. In section 2.4, we have maintained that the noun *intelligence* was the most widely used in 1,000 abstracts, followed by the word *AI*, the noun *study*, the noun *technology*, and the noun *result*, in descending order. Finally, we have contended that the word *AI* is the highest in the closeness centrality. To be more specific, it showed up in the center of the map, thereby implying that it is the most centric and central.

References

[1] Kang, N. (2023a). K-Pop in BBC News: A Big Data Analysis. *Advances in Social Sciences Research Journal*, *10*(2), 156-169.

[2] Kang, N. (2023b). K-Dramas in Google: A NetMiner Analysis. *Transaction on Engineering and Computing Sciences*, *11*(1), 193-216.

[3] Kang, N. (2023c). A Comparative Analysis of Tolerate and Put up with in the COCA. *Semiconductor and optoelectronics* 42(1): 1468-1476.

[4] Kang, N. (2023d). Sure of and Sure about in Corpora and ChatGPT. *Journal of Harbin Engineering University* 44(7): 1347-1351.

[5] Kang, N. (2023e). Turn out adj and Turn out to be adj in the Now Corpus and ChatGPT. *Journal of Harbin Engineering University* 44(8): 825-831.

[6] Kang, N. (2023f). Care for and Like in Corpora and ChatGPT. *Semiconductor and optoelectronics* 42(2): 188-198.