

A Study on Method of Trend Researches

: Focused on Structures of the Academic World
in Science Education

Seok-Hyun Ga*, Eunjin Jang, Chan-Jong Kim, Seung-Urn Choe+
(Seoul National University)



Why is it important to comprehend research trends?

- Trends of research in academic publications represent research interests, approaches, as well as the shared knowledge of researchers (Lin et al., 2014).
- Periodic reports of research trends can help science education researchers **understand the recent status in this research field** (Lin et al., 2014).
- Revealing research trends contribute to develop a new research topics through identifying previously non-researched areas.

Current Studies of Trends Research in Science Education

Journal	Authors	Target Journal	N	Done by	Focus
IJSE	Tsai et al(2005) Taiwan	IJSE, JRST, SE (2008-2012)	869	Human	Frequency of - country - research type - research topic
IJSE	Lee et al(2009) Taiwan	IJSE, JRST, SE (2003-2007)	802	Human	Frequency of - country - research type - research topic - citing
IJSE	Lin et al(2014) Taiwan	IJSE, JRST, SE (1998-2002)	900	Human	Frequency of - country - research type - research topic - citing - author
JSET	Chang et al(2010) Taiwan	IJSE, JRST, RSE, SE	1401	Computer	Frequency of - research topic - country - author - citing

Limitation of Current Studies

- **Only focused on frequency**
 - All of researches were focused on “FREQUENCY” of
 - country, research type, research topic, the number of citing, author, and so on.
 - These researches could not provide correlations between numerous keywords.
- **Too much labor done by human**
 - Most of researches of trends were performed by human.
 - Due to human limit, the number of analyzed articles was small (under 1,000)

The Purpose of Research

- Developing the methodology to calculating the correlations of research **keywords** between journal articles.
- Applying the methodology to the field of science education, and examining its results

Research Design

Data Mining

- Transforming article information form unstructured to structured format (from HTML to Excel)

Data Analysis

- Determine the unit of analysis
- Develop the automatic analysis program using R

Data Visualization

- Determine the format of graph which can visualize effectively
- Visualize information using network graph

Method: (1) Data Mining

- Transforming unstructured data into structured data through HTML Parsing

The screenshot shows the journal's website interface. The main article title is "The role of perspective taking in how children connect reference frames when explaining astronomical phenomena" by Julia D. Plummer, Corinne A. Bower, & Lynn S. Liben. The page includes navigation menus, a sidebar with "Most read" and "Most cited" articles, and a main content area with the article title, authors, and publication details. There are also buttons for "Full text HTML", "PDF", and "Supplemental".

The screenshot shows the HTML source code of the article page. The code is color-coded to show different elements: blue for tags, red for attributes, green for text, and purple for links. The code includes a title tag, a h2 tag for the article title, and a h3 tag for the subtitle. It also includes a DOI link and a link to the full text HTML. The code is as follows:

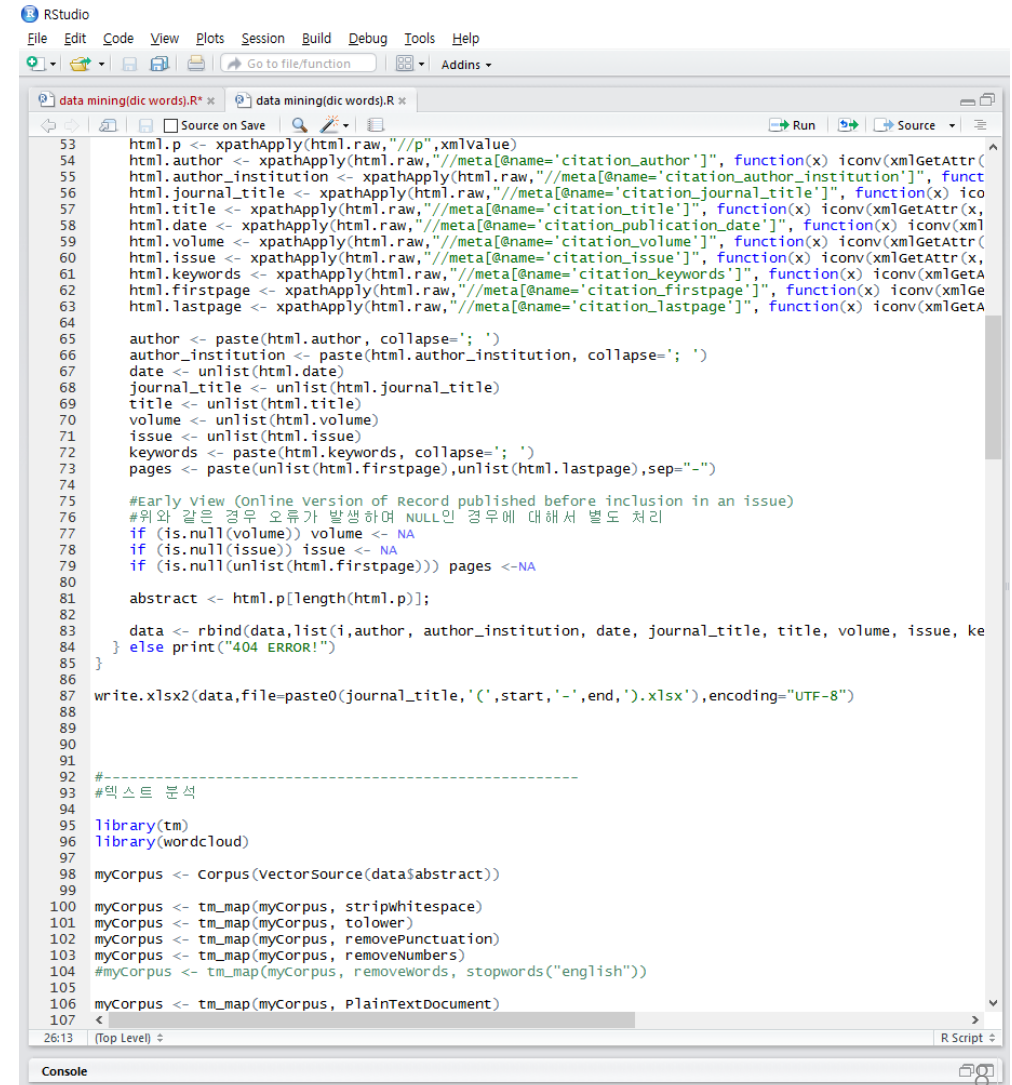
```
<!--totalCount9--><!--modified:1458727172000--><!--ARTICLE: primary_article-->
<div class="article original as2"><h2 class="clear toheading"><span class="
sub-group">Articles </span></h2><div class="gutter"><div class="field
script_only"><input type="checkbox" name="doi" value="
10.1080/09500693.2016.1140921" id="10.1080/09500693.2016.1140921" />
</div><label for="10.1080/09500693.2016.1140921"><a class="entryTitle" href=
"/doi/full/10.1080/09500693.2016.1140921#abstract"><h3>The role of
perspective taking in how children connect reference frames when explaining
astronomical phenomena</h3></a></label><span class="subtitle"
></span><h4><span class="hFld-ContribAuthor"><a href=
"/author/Plummer%2C+Julia+D">Julia D. Plummer</a></span>, <span class=
"/author/Bower%2C+Corinne+A">Corinne A. Bower</a></span> &amp; <span class="hFld-ContribAuthor"><a href=
"/author/Liben%2C+Lynn+S">Lynn S. Liben</a></span><br />pages 345-365
</h4><div class="access accessmodule access_no"><a class="txt" href=
"/doi/full/10.1080/09500693.2016.1140921#abstract">Full text HTML</a><a class
="pdf" target="_blank" href="/doi/pdf/10.1080/09500693.2016.1140921">PDF
</a><a class="suppl" href=
"/doi/suppl/10.1080/09500693.2016.1140921#tabModule">Supplemental</a><a
class="last accessIconLink" target="_blank" href=
"/doi/pdf/10.1080/09500693.2016.1140921"><div class="accessIconWrapper
access_no accessIcon">Access options</div></a></div><ul class="clear
doimetalist"><li><strong>DOI:</strong><strong>10.1080/09500693.2016.1140921
</li><li><strong>Published online:</strong> 14 Mar 2016
</li><li><strong>Citing articles:</strong> 0</li></ul></div></div></div></div>
```

The screenshot shows an Excel spreadsheet with a table of article data. The table has columns for author, journal, title, volume, issue, keyword, and pages. The data is as follows:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1		i	author	author	date	Journal	title	volume	issue	keyword	pages	abstract		
2	505	21128	Ryoo, Kihy	University	2014/02/0	Journal of Designing	51	2	generating	147-174	We compared designs of guidance t			
3	506	21131	Wang, Jinq	National F	2014/02/0	Journal of Exploring	51	2	comprehe	175-191	This study used a Chinese-language			
4	507	21132	Oliveira, A	State Univ	2014/02/0	Journal of Death disc	51	2	death inst	117-146	Little research has been conducted c			
5	508	21133	Mendonça	Universidã	2014/02/0	Journal of An instrun	51	2	argument	192-218	Previous research on argumentation			
6	509	21135	Copur-Gei	University	2014/02/0	Journal of The effect	51	2	reform-ori	219-249	This study examines the impact of a			
7	511	21127	Birmingham	Loyola Un	2014/03/0	Journal of Putting or	51	3	civic actio	286-314	Socioscientific issues in connection t			
8	512	21129	Kisiel, Jam	California	2014/03/0	Journal of Clarifying	51	3	informal s	342-367	If we truly wish to promote science e			
9	513	21130	Feinstein, U	University	2014/03/0	Journal of Scien	51	3	informal;	368-394	Science museums and science cente			
10	514	21134	Bricker, Le	University	2014/03/0	Journal of "What cor	51	3	culture an	260-285	In this paper, we explore the details			
11	515	21141	Rahm, Jr	Université	2014/03/0	Journal of Reframing	51	3	informal s	395-406	Informal science education is a broa			
12	517	21144	Polman, Jc	University	2014/03/0	Journal of Science ne	51	3	science lit	315-341	This paper explores how participatin			
13	519	21136	Fortus, Da	Weizmann	2014/04/0	Journal of Measuring	51	4	motivator	497-522	Continuing motivation for science le			
14	520	21137	Tal, Tali; L	Technion,	2014/04/0	Journal of Exemplary	51	4	field trips	430-461	In light of the literature that deals w			
15	521	21138	Donovan, S	Stanford L	2014/04/0	Journal of Playing wi	51	4	science lit	462-496	Race has been a longstanding topic			
16	522	21143	Zhai, Junq	Zhejiang (2014/04/0	Journal of Communi	51	4	school vis	407-429	Botanic gardens are popular destina			
17	523	21145	DeBoer, G	AAAS Proj	2014/04/0	Journal of Comparin	51	4	assessment	523-554	Online testing holds much promise			
18	524	21116			2014/05/0	Journal of Issue Infor	51	5	fmi-fmv	(2014),	Issue Information. J. Res. Sci.			
19	525	21140	Patchen, T	California	2014/05/0	Journal of Diversifyin	51	5	authority;	606-634	Recent calls asking science teachers			
20	526	21146	Chang, Mi	University	2014/05/0	Journal of What mat	51	5	race; unde	555-580	This longitudinal study examined fa			
21	527	21147	Lee, Hee-S	University	2014/05/0	Journal of Assessmer	51	5	scientific	581-605	Though addressing sources of uncer			
22	528	21148	Diamond, U	University	2014/05/0	Journal of Effectiven	51	5	science co	635-658	Teacher knowledge of science conte			
23	529	21149	Yadav, Am	Purdue Un	2014/05/0	Journal of Case-base	51	5	case-base	659-677	Recently, there has been a push wit			
24	530	21117			2014/08/0	Journal of Issue Infor	51	6	fmi-fmvii	(2014),	Issue Information. J. Res. Sci.			

Method: (2) Data Analysis

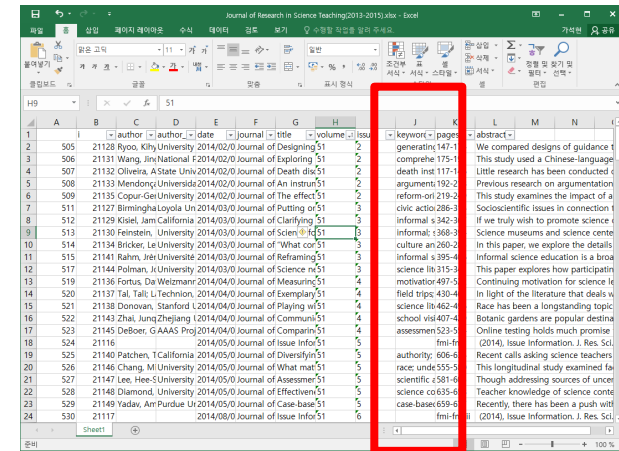
- Designing analysis algorithm
- Developing computer program using R
- Developed computer program process the data which is already mined.



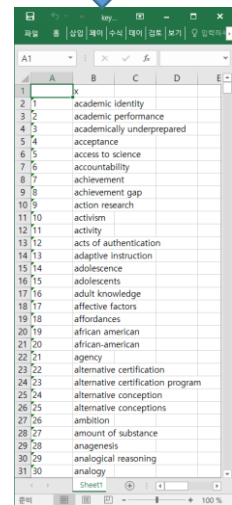
```
RStudio
File Edit Code View Plots Session Build Debug Tools Help
Go to file/function Addins
data mining(dic words).R* data mining(dic words).R
Source on Save Run Source
53 html.p <- xpathApply(html.raw,"//p",xmlvalue)
54 html.author <- xpathApply(html.raw,"//meta[@name='citation_author']", function(x) iconv(xmlGetAttr(
55 html.author_institution <- xpathApply(html.raw,"//meta[@name='citation_author_institution']", funct
56 html.journal_title <- xpathApply(html.raw,"//meta[@name='citation_journal_title']", function(x) ico
57 html.title <- xpathApply(html.raw,"//meta[@name='citation_title']", function(x) iconv(xmlGetAttr(x,
58 html.date <- xpathApply(html.raw,"//meta[@name='citation_publication_date']", function(x) iconv(xml
59 html.volume <- xpathApply(html.raw,"//meta[@name='citation_volume']", function(x) iconv(xmlGetAttr(
60 html.issue <- xpathApply(html.raw,"//meta[@name='citation_issue']", function(x) iconv(xmlGetAttr(x,
61 html.keywords <- xpathApply(html.raw,"//meta[@name='citation_keywords']", function(x) iconv(xmlGETA
62 html.firstpage <- xpathApply(html.raw,"//meta[@name='citation_firstpage']", function(x) iconv(xmlGe
63 html.lastpage <- xpathApply(html.raw,"//meta[@name='citation_lastpage']", function(x) iconv(xmlGETA
64
65 author <- paste(html.author, collapse='; ')
66 author_institution <- paste(html.author_institution, collapse='; ')
67 date <- unlist(html.date)
68 journal_title <- unlist(html.journal_title)
69 title <- unlist(html.title)
70 volume <- unlist(html.volume)
71 issue <- unlist(html.issue)
72 keywords <- paste(html.keywords, collapse='; ')
73 pages <- paste(unlist(html.firstpage),unlist(html.lastpage),sep="-")
74
75 #Early View (online version of Record published before inclusion in an issue)
76 #위와 같은 경우 오류가 발생하여 NULL인 경우에 대해서 별도 처리
77 if (is.null(volume)) volume <- NA
78 if (is.null(issue)) issue <- NA
79 if (is.null(unlist(html.firstpage))) pages <-NA
80
81 abstract <- html.p[length(html.p)];
82
83 data <- rbind(data,list(i,author, author_institution, date, journal_title, title, volume, issue, ke
84 } else print("404 ERROR!")
85 }
86
87 write.xlsx2(data,file=paste0(journal_title,'(',start,'-',end,').xlsx'),encoding="UTF-8")
88
89
90
91
92 -----
93 #텍스트 분석
94
95 library(tm)
96 library(wordcloud)
97
98 myCorpus <- Corpus(VectorSource(data$abstract))
99
100 myCorpus <- tm_map(myCorpus, stripwhitespace)
101 myCorpus <- tm_map(myCorpus, tolower)
102 myCorpus <- tm_map(myCorpus, removePunctuation)
103 myCorpus <- tm_map(myCorpus, removeNumbers)
104 #myCorpus <- tm_map(myCorpus, removewords, stopwords("english"))
105
106 myCorpus <- tm_map(myCorpus, PlainTextDocument)
107 <
26:13 (Top Level) R Script
Console
```


Method: (2) Data Analysis

- Assumed that keywords in research article has a meaning of research topic.
- Collecting keywords from mined data to use keywords as the unit of analysis
- Eliminating the duplicated keywords from the list.



Eliminating Duplicated keywords



Method: (2) Data Analysis

- Calculating relations between keywords

article 1	article 2	article 3
<ul style="list-style-type: none">• action research• authentic inquiry• curriculum• teacher cognition	<ul style="list-style-type: none">• constructivism• citizen science• action research• authentic inquiry	<ul style="list-style-type: none">• constructivism• feminism• teacher education• authentic inquiry

- If there are two words(action research, authentic inquiry) in both article 1 and article 2, these keywords is added 2-point correlation.
- If there aren't two words(curriculum, constructivism) in one article at the same time, these keywords isn't added correlation point.

Method: (2) Data Analysis

- Calculating relations between keywords

article 1	article 2	article 3
<ul style="list-style-type: none">• action research• authentic inquiry• curriculum• teacher cognition	<ul style="list-style-type: none">• constructivism• citizen science• action research• authentic inquiry	<ul style="list-style-type: none">• constructivism• feminism• teacher education• authentic inquiry

- If there are two words(action research, authentic inquiry) in both article 1 and article 2, these keywords is added 2-point correlation.
- If there aren't two words(curriculum, constructivism) in one article at the same time, these keywords isn't added correlation point.

Method: (2) Data Analysis

- Calculating relations between keywords

article 1	article 2	article 3
<ul style="list-style-type: none">• action research• authentic inquiry• curriculum• teacher cognition	<ul style="list-style-type: none">• constructivism• citizen science• action research• authentic inquiry	<ul style="list-style-type: none">• constructivism• feminism• teacher education• authentic inquiry

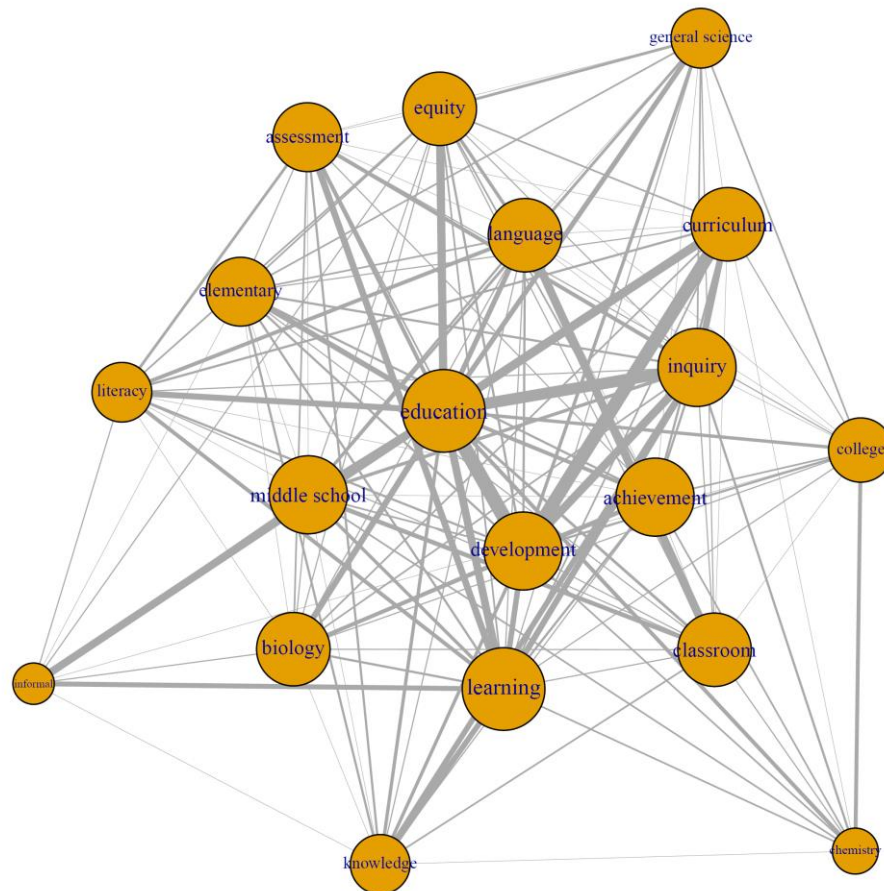
- If there are two words(action research, authentic inquiry) in both article 1 and article 2, these keywords is added 2-point correlation.
- If there aren't two words(curriculum, constructivism) in one article at the same time, these keywords isn't added correlation point.

Method: (3) Data Visualization

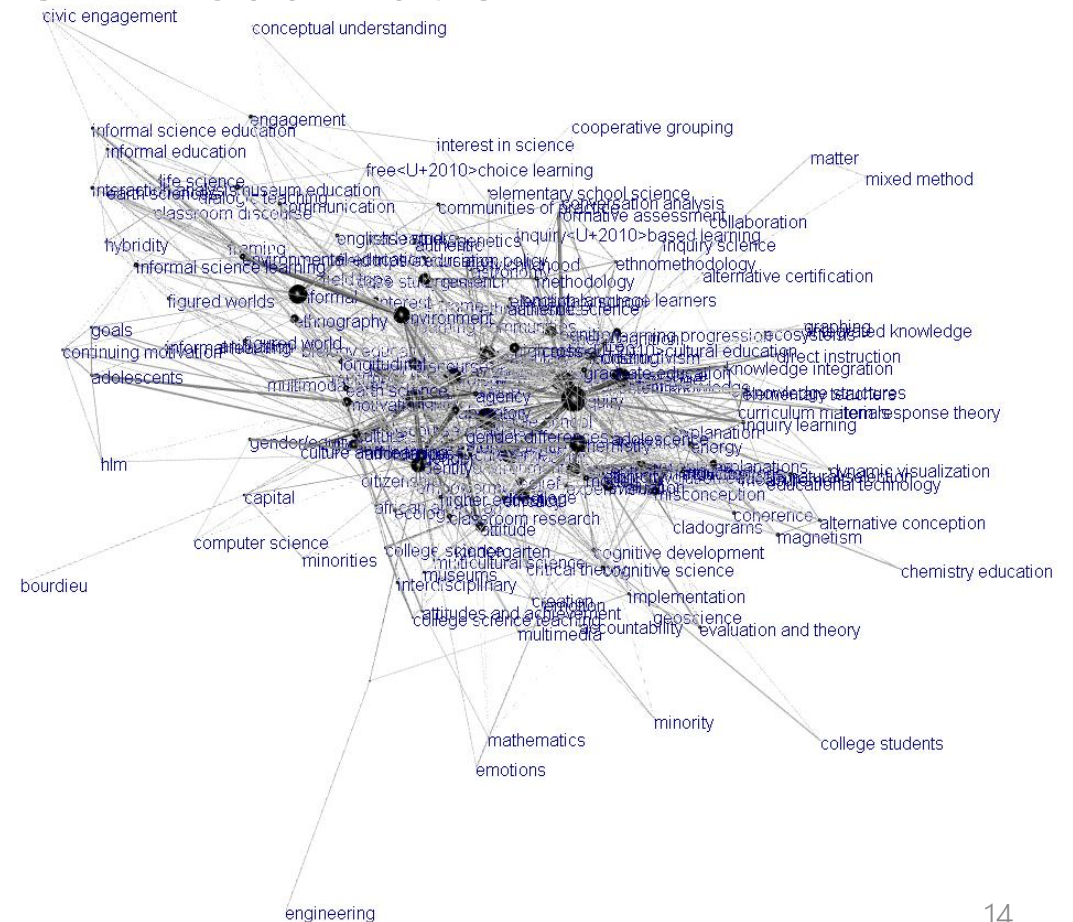
- Visualizing data through network graph to show correlation between keywords effectively.
 - A **node** is a keyword which means "topic" or "terms", and the **size of node** is frequency of appearance.
 - A **edge** is correlation between two nodes, and the **width of edge** is degree of correlation.

Method: (3) Data Visualization

- 2D Visualization



- 3D Visualization



Method: (3) Data Visualization

- Eliminating general keywords(node)
 - *General Keyword: the keyword which is commonly used in science education, but doesn't have meaningful information about research topic



Method: (3) Data Visualization

- Criteria of eliminating node
 - Keywords that mostly used or nearly used are not meaningful, so these keywords are eliminated automatically by computer program. (Eliminate the keyword which appear with a probability of 0.02)

$$0.02 < P(K_i)$$

- Two keywords that has high correlation between them are integrated into one node. (Integrate the keywords which have a 0.90 correlation)

$$P(K_j|K_i) < 0.90$$

- Limitation: The node is mechanically eliminated or integrated, so it caused inaccuracy.

* $P(K_i)$: Apperence Probability of i – th keyword

Method: (3) Data Visualization

- Done by Researcher
 - Most frequently used keywords – i.e. education, science, development – can not provide meaningful information about research topics, so these keyword are selectively eliminated.

education, learning, development, curriculum development

- Similar keywords are integrated into one node.

curriculum, curriculum development / learning progression, learning progressions / Informal, informal science / argument, argumentation / attitude, attitudes / middle school, middle school science

- Limitation: It depends on subjectivity of researcher.

Results

- Analysis result of IJSE or JRST articles from 2008 to 2015

IJSE

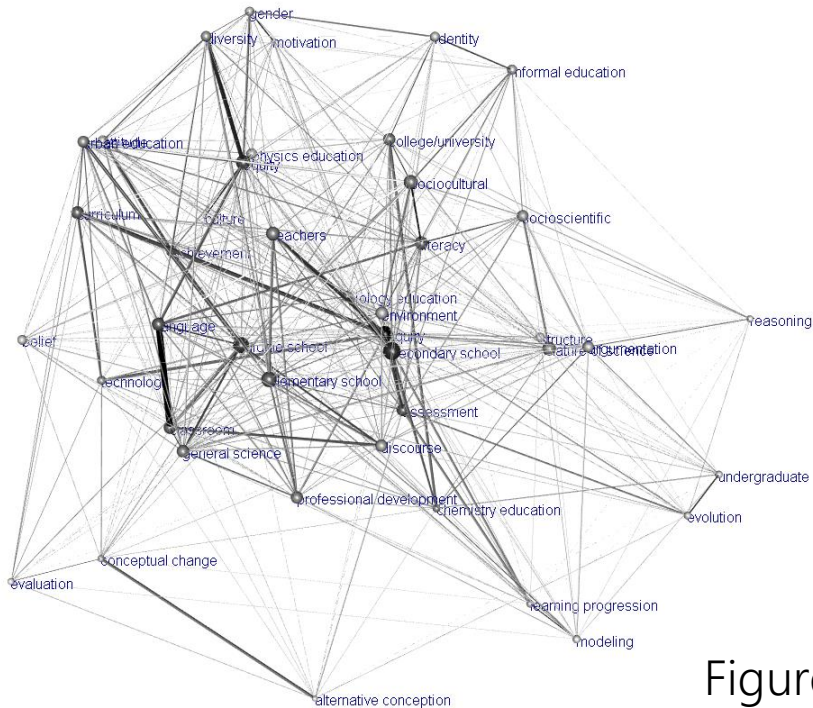


Figure 4

JRST

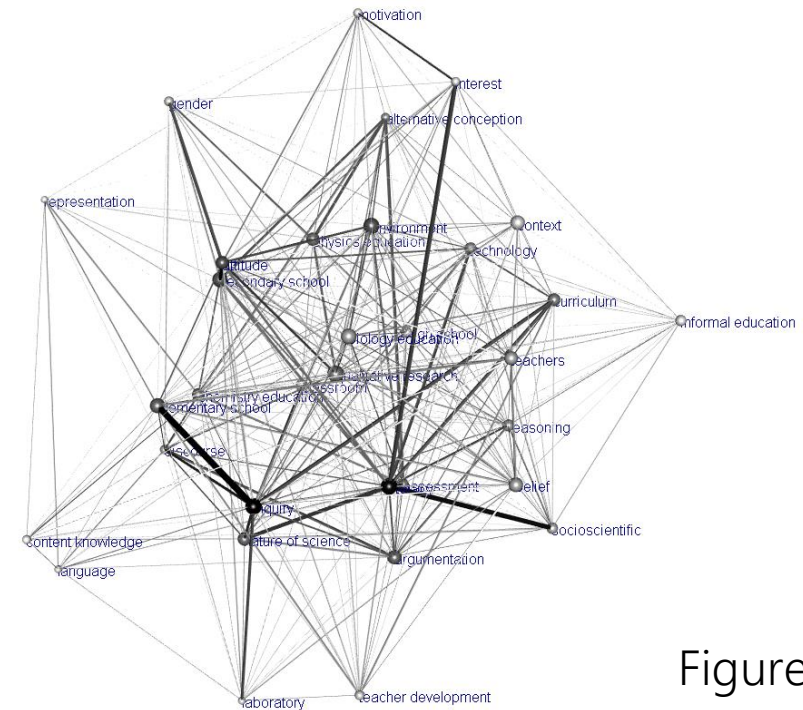


Figure 5

- There are no significant difference between results of JRST and IJSE.

Conclusion

- Analysis of research trends can be processed by computer.
 - So, huge size of articles can be analyzed automatically.
- Using this methodology in science education,
 - **Structure of research field** in science education can be visualized.
 - **The most prominent keywords** and **high correlation keywords** can be revealed.

Limitations of Research

- Some of journals or past journals(before 2007) **doesn't have keywords** in their articles, so these journals cannot apply this keyword-based methodology.
 - In order to overcome these limitation, new methodology developed that "authentic keywords" are extracted from full-text of articles.
 - When it is done, there will be **significant difference between analysis graphs of 4 years**, and **forecasting research trends** can be possible.

Limitations of Research

- This research help researchers who want to know about research trends.
- But it is wondered how this research might contribute to teaching and learning.

Thank you!

If you want to get computer programming source codes of this research, please contact me by email (shga@snu.ac.kr)