

연구동향 분석법에 대한 방법론적 제안

: 과학교육 분야 연구 세계 구조의 가시화

가석현*, 장은진, 김찬종, 최승언+
(서울대학교)



연구동향을 살피는 것이 왜 중요한가?

- 학술지에 실린 논문들의 연구 동향을 살피는 것은, 최근 학계에
서 관심을 가지는 연구 분야나 접근법들을 알 수 있을 뿐만 아니
라, 연구자들 간의 공유된 지식을 알아낼 수 있다. (Lin et al.,
2014).
- 정기적인 연구동향 연구들은 과학교육학자들이 해당 연구 분야
의 최근 동향을 알 수 있도록 돕는다 (Lin et al., 2014).
- 연구동향을 규명하는 것은 아직까지 연구되지 않은 영역을 발견
하여, 새로운 연구주제를 찾을 수 있도록 돕는다.

선행 연구 과학교육 분야 연구 동향

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

선행 연구들의 한계점

- **빈도 분석에 한정됨**

- 대부분의 연구들이 각 논문에 대한 여러 속성들을 빈도분석
 - 국가, 연구 종류, 연구 주제, 인용수, 저자 등.
- 이러한 연구들은 여러 연구들 간의 상관 관계를 규명하는데 한계가 존재

- **수작업에 의해 분석**

- 대부분의 연구들은 사람의 수작업에 의해 분류되거나 분석됨
- 사람이 수행할 수 있는 데이터 양의 한계로 인해 분석된 논문의 수가 많지 않으며, 여러 연구들 간의 상관을 밝히는 것은 거의 불가능함.

연구 목적

- 학술지 논문들의 키워드들을 이용하여 여러 연구 주제들 간의 상관 관계를 규명할 수 있는 "방법론"을 개발
- 해당 "방법론"을 컴퓨터 프로그래밍으로 구현
- 과학교육 연구 분야에 적용

연구 설계

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

연구 방법: (1) Data Mining

- HTML 파싱을 이용하여 비정형화된 데이터들을 정형화된 데이터 (ex. Excel, MySQL 등)로 변환

The screenshot shows the website for the International Journal of Science Education, Volume 38, Issue 3, 2016. The main article displayed 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 links like "View all volumes and issues", "Current issue", and "Latest articles". It also features a "Publication History" section and "Access options" for the full text HTML and PDF.

The screenshot shows the HTML source code of the article page. The code is color-coded and includes various HTML tags and attributes. Key elements include the article title, authors, and a DOI link. The code is as follows:

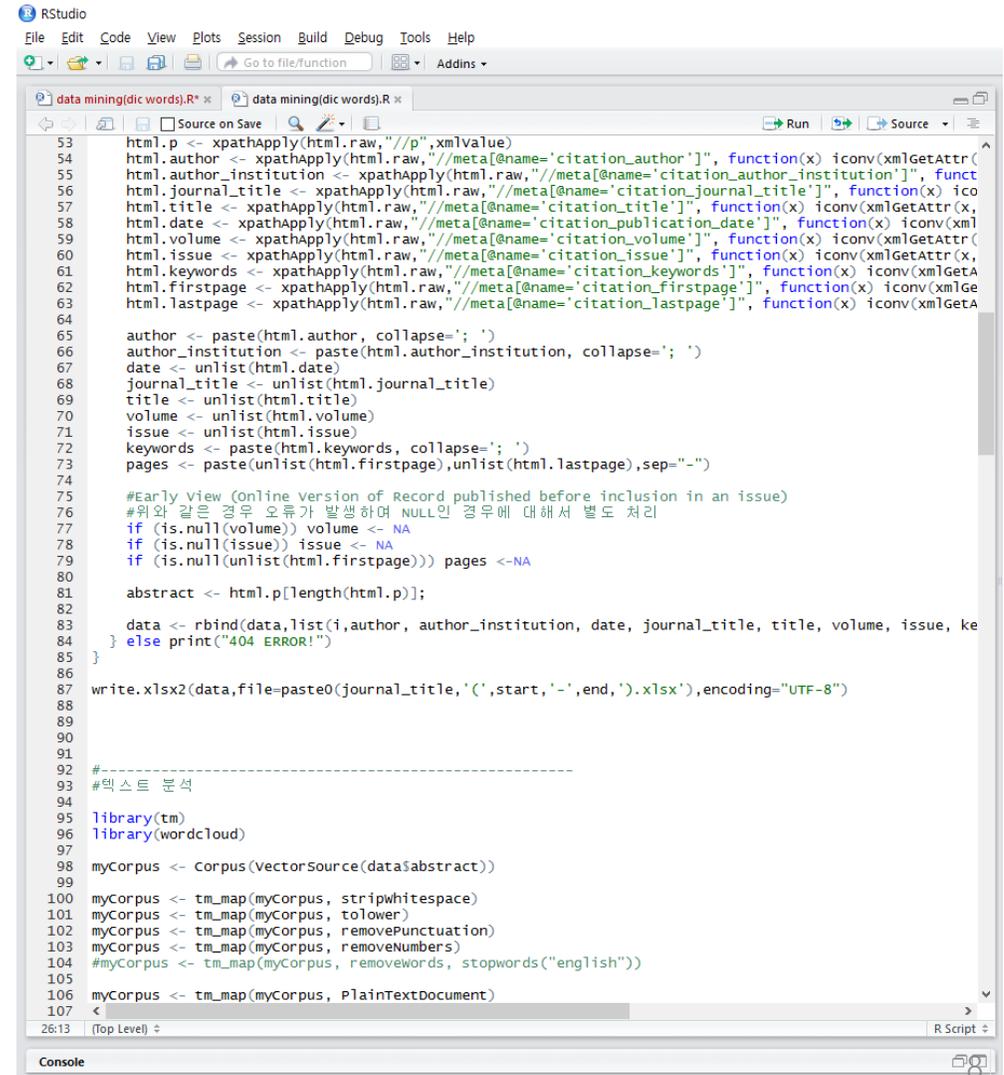
```
<!--totalCount9--><!--modified:1458727172000--><!--ARTICLE: primary_article-->
<div class="article original as2"><h2 class="clear toheading"><span class="
sub-group">Articles </span></h2><div class="gutter"><div class="field
script_only"><input type="checkbox" name="doi" value="
"10.1080/09500693.2016.1140921" 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=
"hFld-ContribAuthor"><a href="/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 class=
```

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	niversity	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, Jrêr	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;	Li 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	tanford L	2014/04/0	Journal of Playing w	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	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	niversity	2014/05/0	Journal of Effectiven	51	5	science co	635-658	Teacher knowledge of science conte			
23	529	21149	Yadav, Am	Purdue Ur	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.			

연구 방법: (2) Data Analysis

- 알고리즘 설계
- R을 이용하여 프로그래밍
- 수집된 데이터를 분석



```
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 발생하며 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 # 텍스트 분석
93
94 library(tm)
95 library(wordcloud)
96
97 myCorpus <- Corpus(VectorSource(data$abstract))
98
99 myCorpus <- tm_map(myCorpus, stripwhitespace)
100 myCorpus <- tm_map(myCorpus, tolower)
101 myCorpus <- tm_map(myCorpus, removePunctuation)
102 myCorpus <- tm_map(myCorpus, removeNumbers)
103 myCorpus <- tm_map(myCorpus, removeWords, stopwords("english"))
104
105 myCorpus <- tm_map(myCorpus, PlainTextDocument)
106
107 <
26:13 (Top Level) R Script
Console
```

연구 방법: (2) Data Analysis

- 논문의 키워드는 해당 논문의 핵심 주제를 담고 있다고 가정
- 키워드를 하나의 분석 단위로 설정
- 모든 논문의 키워드를 DB화 한 후, 중복된 키워드를 제거

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
		author	author	date	journal	title	volume	issue	keyword	pages	abstract			
1														
2	505	21128	Ryoo, Kihy	2014/02/0	Journal of Designing	51	2	2	generating	147-1	We compared designs of guidance t			
3	506	21131	Wang, Jinx	2014/02/0	Journal of Exploring	51	2	2	comprehe	175-1	This study used a Chinese-language			
4	507	21132	Oliveira, A	2014/02/0	Journal of Death	51	2	2	death inl	117-1	Little research has been conducted i			
5	508	21133	Mendonça	2014/02/0	Journal of An Instrum	51	2	2	argument	192-2	Previous research on argumentation			
6	509	21135	Copur-Gei	2014/02/0	Journal of The effect	51	2	2	reform-or	219-2	This study examines the impact of a			
7	511	21127	Birmingham	2014/03/0	Journal of Putting or	51	3	3	civic actio	286-3	Socioscientific issues in connection t			
8	512	21129	Kisei, Jam	2014/03/0	Journal of Clarifyin	51	3	3	informal s	342-9	If we truly wish to promote science i			
9	513	21130	Feinstein, L	2014/03/0	Journal of Science	51	3	3	informal s	336-8	Science museums and science cente			
10	514	21134	Bricker, Le	2014/03/0	Journal of "What cor	51	3	3	culture an	260-2	In this paper, we explore the details			
11	515	21141	Rahm, J	2014/03/0	Journal of Reframing	51	3	3	informal s	395-4	Informal science education is a broa			
12	517	21144	Polman, J	2014/03/0	Journal of Science	51	3	3	science lit	315-3	This paper explores how participati			
13	519	21136	Fortus, Da	2014/04/0	Journal of Measurin	51	4	4	Continuing	497-5	Continuing motivation for science le			
14	520	21137	Tal, Tali; L	2014/04/0	Journal of Exemplary	51	4	4	field trips	430-4	In light of the literature that deals vi			
15	521	21138	Donovan, Stan	2014/04/0	Journal of Playing w	51	4	4	science lit	462-4	Race has been a longstanding topic;			
16	522	21143	Zhai, Junq	2014/04/0	Journal of Commun	51	4	4	school vis	407-4	Botanic gardens are popular destina			
17	523	21145	DeBoer, G	2014/04/0	Journal of Compan	51	4	4	assessment	523-5	Online testing holds much promise			
18	524	21116		2014/05/0	Journal of Issue Info	51	5	5	fmi-fr		(2014), Issue Information. J. Res. Sci.			
19	525	21140	Patchen, T	2014/05/0	Journal of Diversifyin	51	5	5	authority;	606-6	Recent calls asking science teachers			
20	526	21146	Chang, M	2014/05/0	Journal of What mat	51	5	5	race: unde	555-5	This longitudinal study examined fa			
21	527	21147	Lee, Hee-S	2014/05/0	Journal of Assessm	51	5	5	scientific	501-6	Though addressing sources of unce			
22	528	21148	Diamond, U	2014/05/0	Journal of Effectiven	51	5	5	science co	635-6	Teacher knowledge of science cont			
23	529	21149	Yadav, An	2014/05/0	Journal of Case-base	51	5	5	case-base	659-6	Recently, there has been a push wit			
24	530	21117		2014/06/0	Journal of Issue Info	51	6	6	fmi-fr		(2014), Issue Information. J. Res. Sci.			

Eliminating Duplicated keywords

	A	B	C	D	E
1					
2	1	academic identity			
3	2	academic performance			
4	3	academically underprepared			
5	4	acceptance			
6	5	access to science			
7	6	accountability			
8	7	achievement			
9	8	achievement gap			
10	9	action research			
11	10	activism			
12	11	activity			
13	12	acts of authentication			
14	13	adaptive instruction			
15	14	adolescence			
16	15	adolescents			
17	16	adult knowledge			
18	17	affective factors			
19	18	affordances			
20	19	african american			
21	20	african-american			
22	21	agency			
23	22	alternative certification			
24	23	alternative certification program			
25	24	alternative conception			
26	25	alternative conceptions			
27	26	ambition			
28	27	amount of substance			
29	28	anagness			
30	29	analogical reasoning			
31	30	analogy			

연구 방법: (2) Data Analysis

- 키워드 간의 상관도를 계산

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

- 두 단어(action research, authentic inquiry)가 모두 등장하는 논문이 2편(article 1, 2) 존재한다면, 이들의 상관도는 "2"
- 두 단어(curriculum, constructivism)가 동시에 등장하는 논문이 단 1편도 없다면 이들의 상관도는 "0"

연구 방법: (2) Data Analysis

- 키워드 간의 상관도를 계산

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

- 두 단어(action research, authentic inquiry)가 모두 등장하는 논문이 2편(article 1, 2) 존재한다면, 이들의 상관도는 "2"
- 두 단어(curriculum, constructivism)가 동시에 등장하는 논문이 단 1편도 없다면 이들의 상관도는 "0"

연구 방법: (2) Data Analysis

- 키워드 간의 상관도를 계산

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

- 두 단어(action research, authentic inquiry)가 모두 등장하는 논문이 2편(article 1, 2) 존재한다면, 이들의 상관도는 "2"
- 두 단어(**curriculum, constructivism**)가 동시에 등장하는 논문이 단 1편도 없다면 이들의 상관도는 "0"

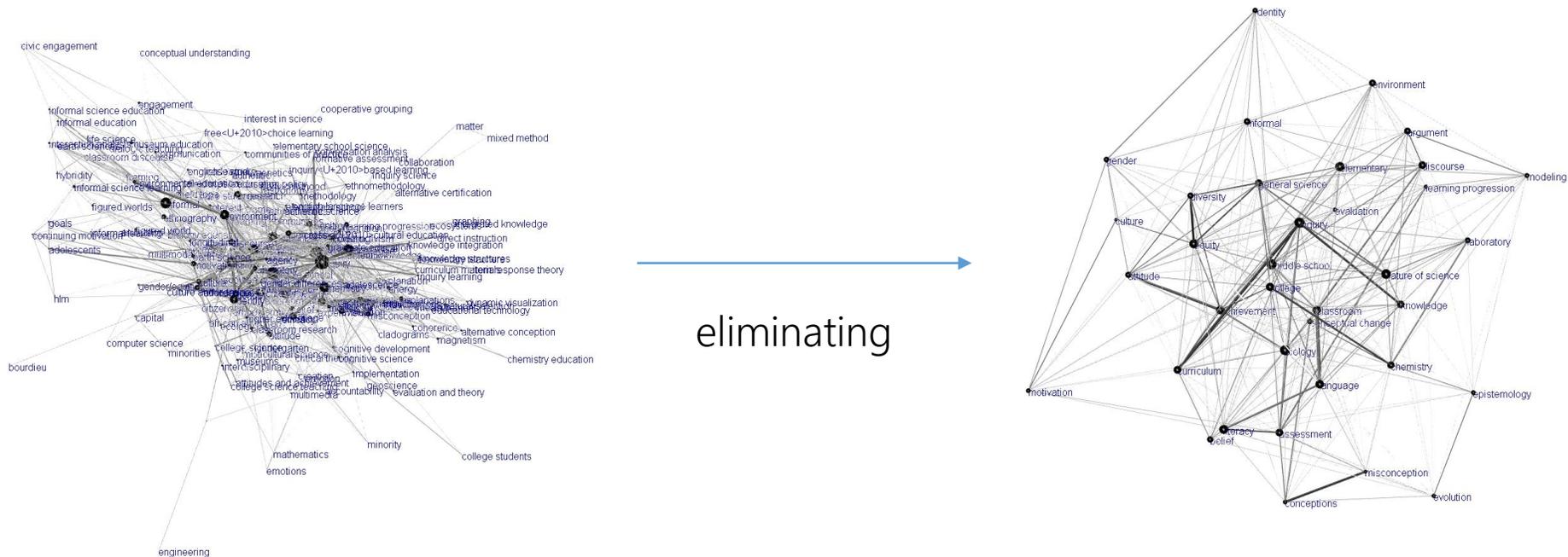
연구 방법: (3) Data Visualization

- 상관도를 효과적으로 가시화 하기 위하여, 네트워크 그래프로 표현
 - 정점(edge) : 하나의 키워드는 하나의 정점으로 시각화. 해당 키워드가 등장하는 빈도를 정점의 크기로 표현
 - 간선(node) : 두 키워드(정점)간의 상관성이 존재함을 나타냄. 상관도를 간선의 굵기로 표현.

연구 방법: (3) Data Visualization

- 일반적인 키워드(정점)를 삭제

* 일반적인 키워드: 과학교육에서 흔하게 등장하는 단어로, 연구의 특질적 요소를 나타내주지 못하는 단어



연구 방법: (3) Data Visualization

- 특별한 기준에 의한 제거

- 거의 등장하지 않는 키워드는 제거 (등장 확률이 3% 미만일 경우 제거)

$$0.03 < P(K_i)$$

- 동시에 등장할 확률이 90%를 넘어설 경우 하나의 정점으로 통합

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

- 한계: 정점의 제거가 기계적으로 이루어지므로, 의미 있는 정점이 누락되는 한계가 존재함.

* $P(K_i)$: Appearance Probability of i - th keyword

Method: (3) Data Visualization

- 연구자에 의한 제거

- 일반적인 키워드(i.e. education, science, development)는 의미있는 정보를 제공하지 못하므로, 선택적으로 제거

education, learning, development, curriculum development

- 비슷한 키워드는 하나의 키워드로 통합

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

- 한계: 연구자의 주관성이 개입됨

연구 결과

- IJSE, JRST의 2008~2015년 투고된 논문을 분석

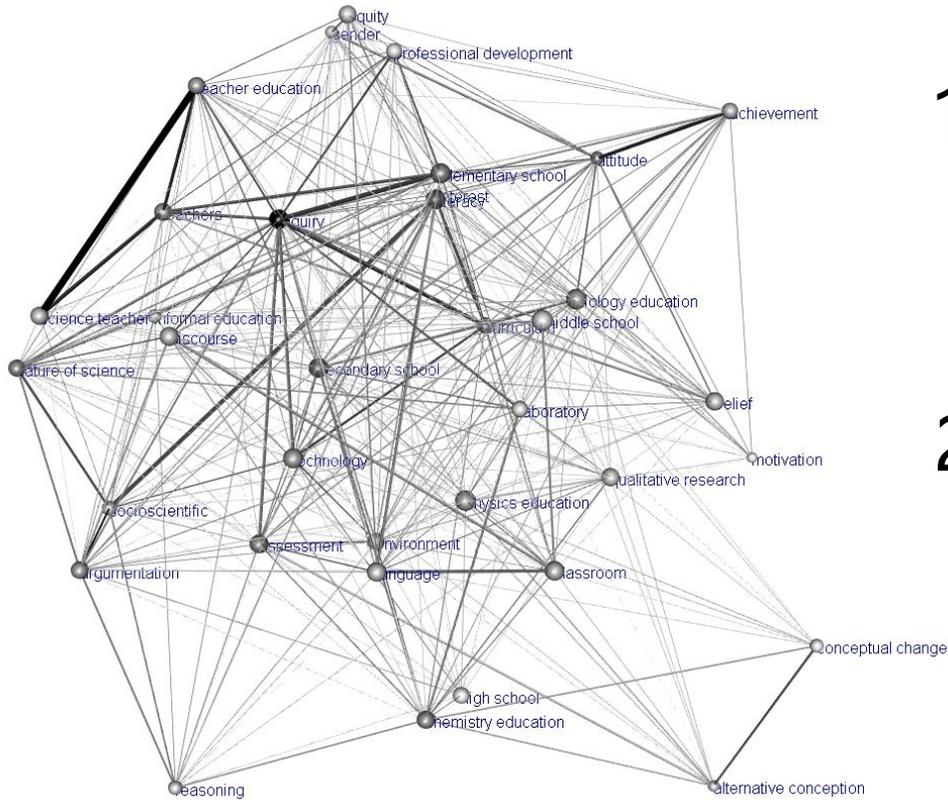


Figure 1

1) 자주 등장하는 키워드

inquiry, secondary school, middle school, language, classroom

2) 상관도가 높은 키워드

language-classroom
inquiry-curriculum
diversity-equity
inquiry-technology

연구 결과

- IJSE, JRST의 2008~2015년 투고된 논문을 논문별로 분석

IJSE

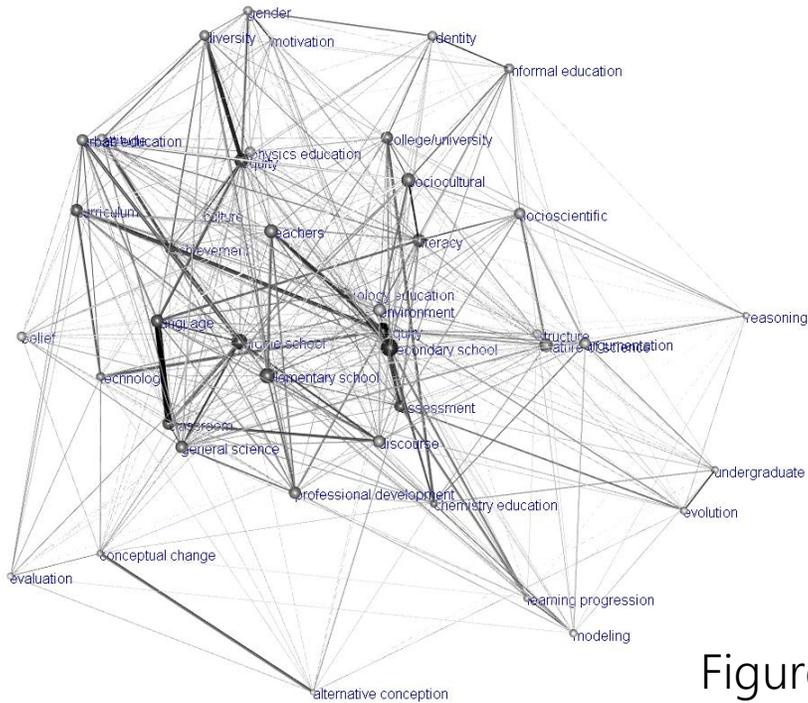


Figure 4

JRST

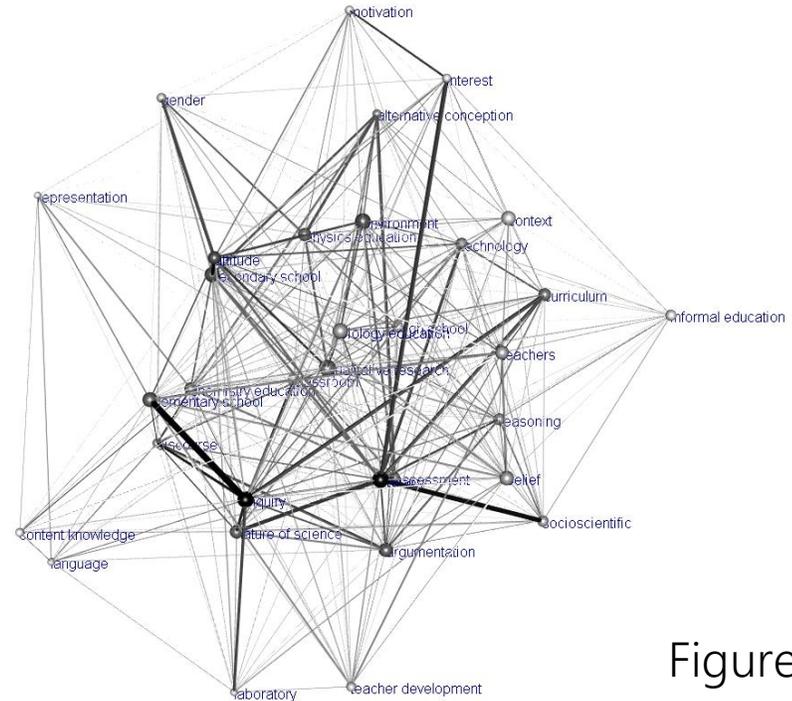


Figure 5

- JRST, IJSE 사이의 큰 차이는 발견되지 않음

결론

- 컴퓨터에 의한 연구 동향 분석의 가능성을 확인
 - 따라서, 매우 많은 수의 논문을 분석하는 것이 가능함
- 과학교육 분야에 이러한 방법을 이용하여,
 - 연구 분야의 구조를 가시화 할 수 있었음
 - 연구 주제들 간의 상관도를 알아볼 수 있었음

연구의 한계

- 일부 학술지나 예전 논문(2007년 이전)은 키워드를 가지고 있지 않아, 키워드 기반의 분석이 불가함
 - 이러한 한계를 극복하기 위해, 논문의 전문을 분석하여 “진정한 키워드”를 찾아내는 방법을 개발할 필요가 있음
 - TF-IDF 알고리즘
- 이것이 가능할 경우, 시간에 따라 변화하는 연구동향을 파악할 수 있고 이를 통해 미래의 연구 동향을 예측할 수 있을 것이라 기대함.

연구의 한계

- 이러한 연구 동향 분석은 자신이 속한 학계의 연구 동향을 알고자 하는 학자들에게 도움이 될 것으로 기대함.
- 하지만, 이러한 연구들이 교사, 학생들의 교수 학습에 어떠한 기여를 할 수 있을 지에 대한 부분이 불명확함.

감사합니다.