

Applications of Multiple and Simple CRA in Text Mining

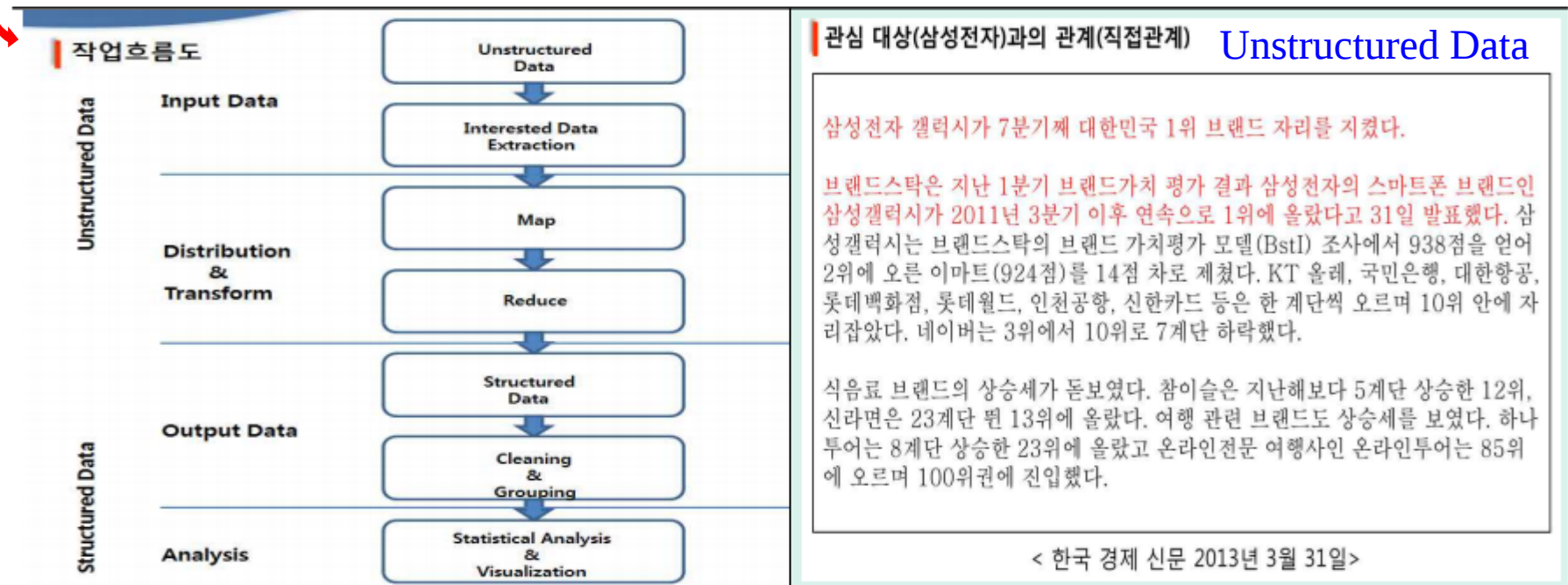
응용 1: 삼성과 애플의 기사의 맵리듀스와 MCA를 활용한 동향 분석

Choi, J. and Choi, Y.S. (2014). Standardizing Unstructured Big Data and Visual Interpretation Using MapReduce and Correspondence Analysis, *The Korean Journal of Applied Statistics*, 27(2), 169-183.

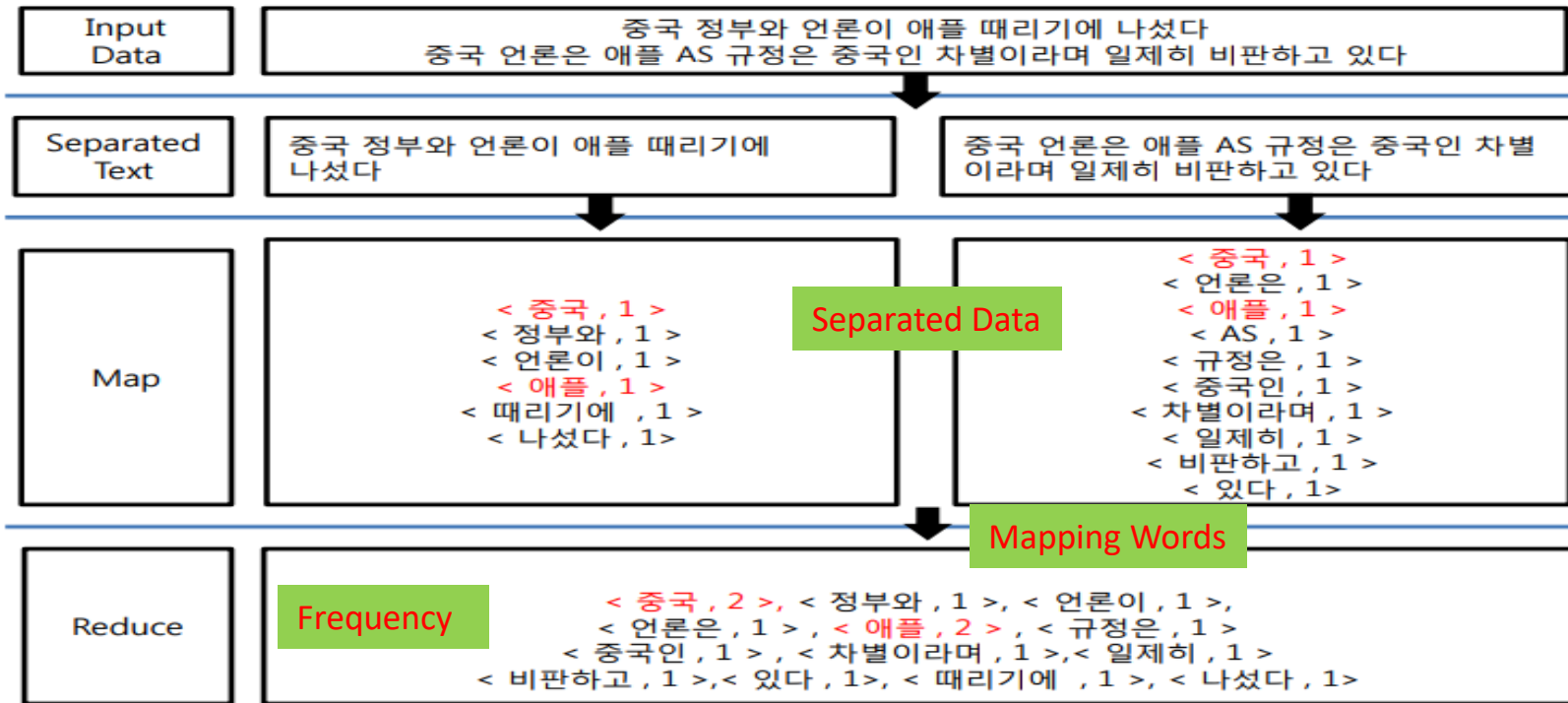
Distribution Processing System : MapReduce Process of Hadoop

Quarter: 1st ~ 4th

1) 단어 세분화 및 군집화 (Structured Data) : 두 기업의 분기별 동향과 이슈(2012.01-2012.12, 한국경제신문)



맵리듀스 과정



삼성전자와 애플의 분기별 동향과 이슈를 살펴봄(한국경제신문2012.01~2012.12).

< 삼성전자와 애플 제목의 분기별 지면 기사 빈도 >

	1분기	2분기	3분기	4분기	합계
삼성전자	61	58	47	40	206
애플	45	43	100	75	263

Top 5% of freq.: 총 150(삼성 70, 애플 105, 삼성&애플 152)

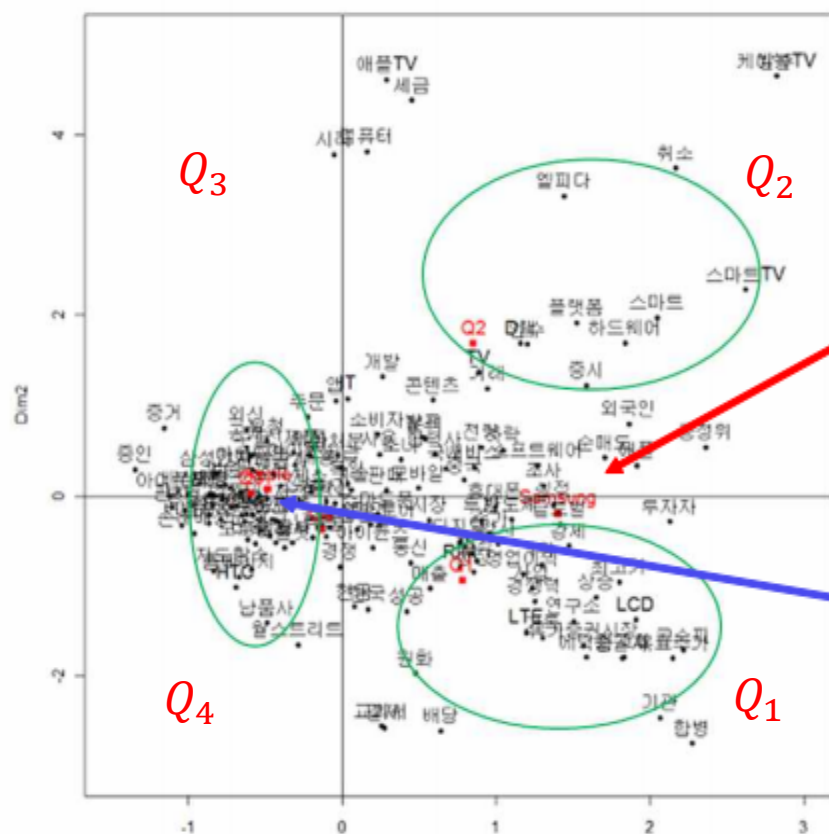
< 맵리듀스 과정을 거친 분기별 직접관계 단어 >

	1분기	2분기	3분기	4분기
삼성전자	2769	2407	2395	2060
애플	2903	2745	5106	4288

Freq. of titles(삼성, 애플)

Total # of words: 24,673

2) 대응분석(Correspondence Analysis)

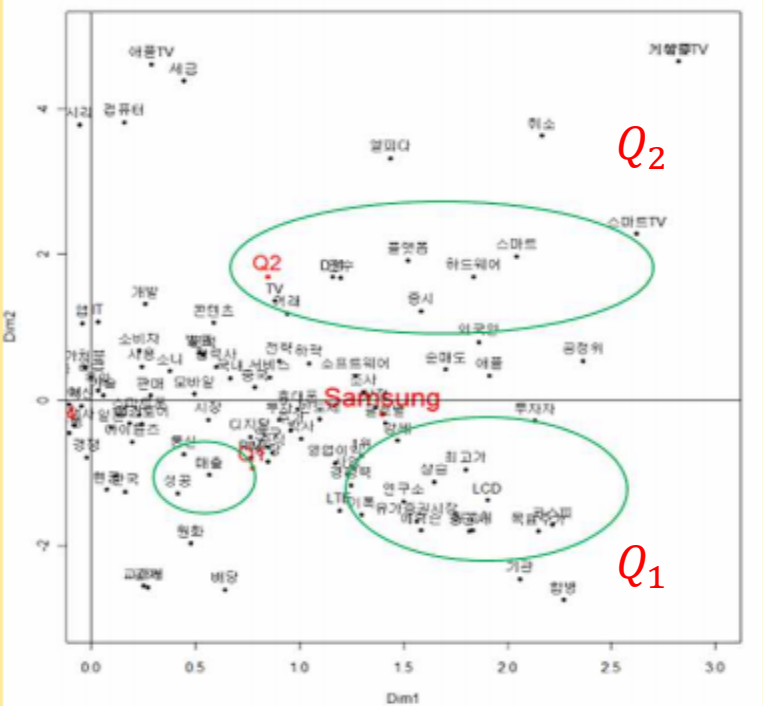


<삼성전자와 애플의 기사¹⁾ 단어의 다중대응분석도>

기업	분기	사분면	단어	의미
Samsung (삼성전자)	Q1(1분기)	제4사분면	매출, 원화, 성공, 상승, 최고가, 영업이익, 1위, 목표주가, 교과서, 경쟁력, 코스피, 한국, 현금 등	성장
	Q2(2분기)	제1사분면	D램, 하드웨어, 스마트, 스마트TV, 애플TV, 콘텐츠, 개발, 모바일, 소프트웨어 등	제품, 사업확장
Apple (애플)	Q3(3분기)	제2사분면	위신, 증거, 증인, 삼성전자, 요청, 제소, 소송, 판매금지, 배심원 등	특히 소송
	Q4(4분기)	제4사분면	지도, 홈페이지, 손해배상, 변호, 제기, 경쟁사, 모토로라, 판결, 경쟁 등	특히 소송

Patent Lawsuit

Expansion



< 제 1사분면과 제 4사분면>

Current Status Analysis

✓ 현황분석 : 2012년

Patent Lawsuit

애플: 특허 관련 소송에 주력

삼성전자 : 사업 확장과 성장에 주력
Business Expansion and Growth

Predictive Analysis

✓ 예측분석 : 2013년 상반기

삼성전자 : 지속적인 성장과 사업 확장 → 상승세 전망

Continuous Growth

-> Prospect for upturn

애플 : 계속되는 특허 관련 소송

-> 슬로건인 혁신적 기업 이미지에 타격

-> 성장의 주춤 전망

Breaking innovation image

숙명의 벽: 삼성의 비창의적

앞으로 삼성이 애플을 넘어서주기를 간절히 바라나 삼성이 수익을 많이 내어도 애플 앞에만 서면 작아진다. 창의적이지 못하기 때문이다. ⇒ 철학/인문/예술/문화 등의 **시선의 높이를 상승**
⇒ **따라 하기를 버리고 창의성/상상력을 지닌 자기 시선으로 세계를 보고 독립적인 삶을 영유**
최진석(2017) 탁월한 사유의 시선- 우리가 꿈꾸는 시대를 위한 철학의 힘

정민지, 최용석(2016) Zonation and Zonation Organism- Understanding and Statistical Applications-

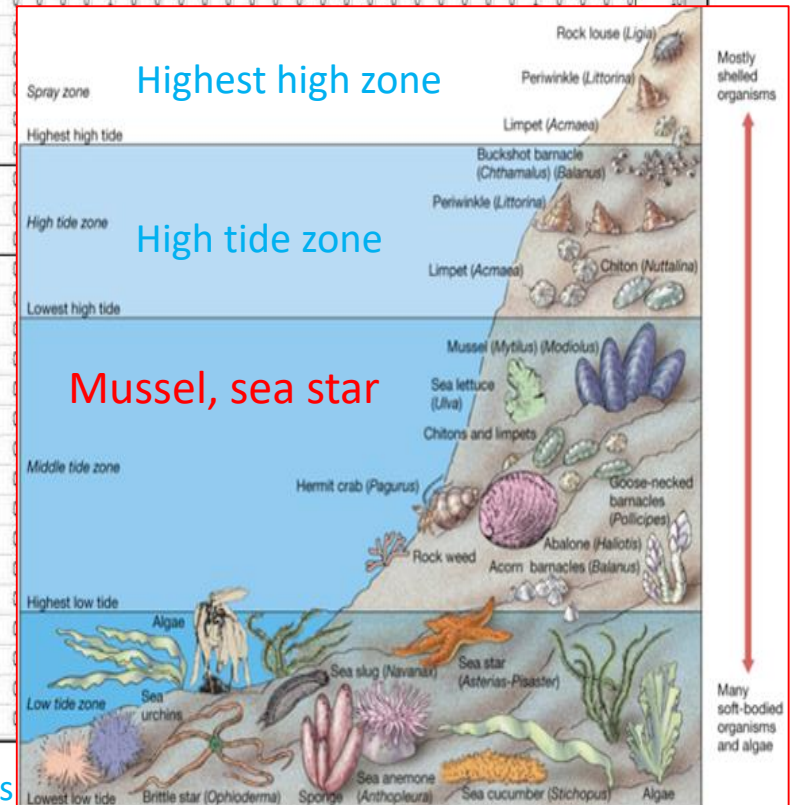
Statistical methods

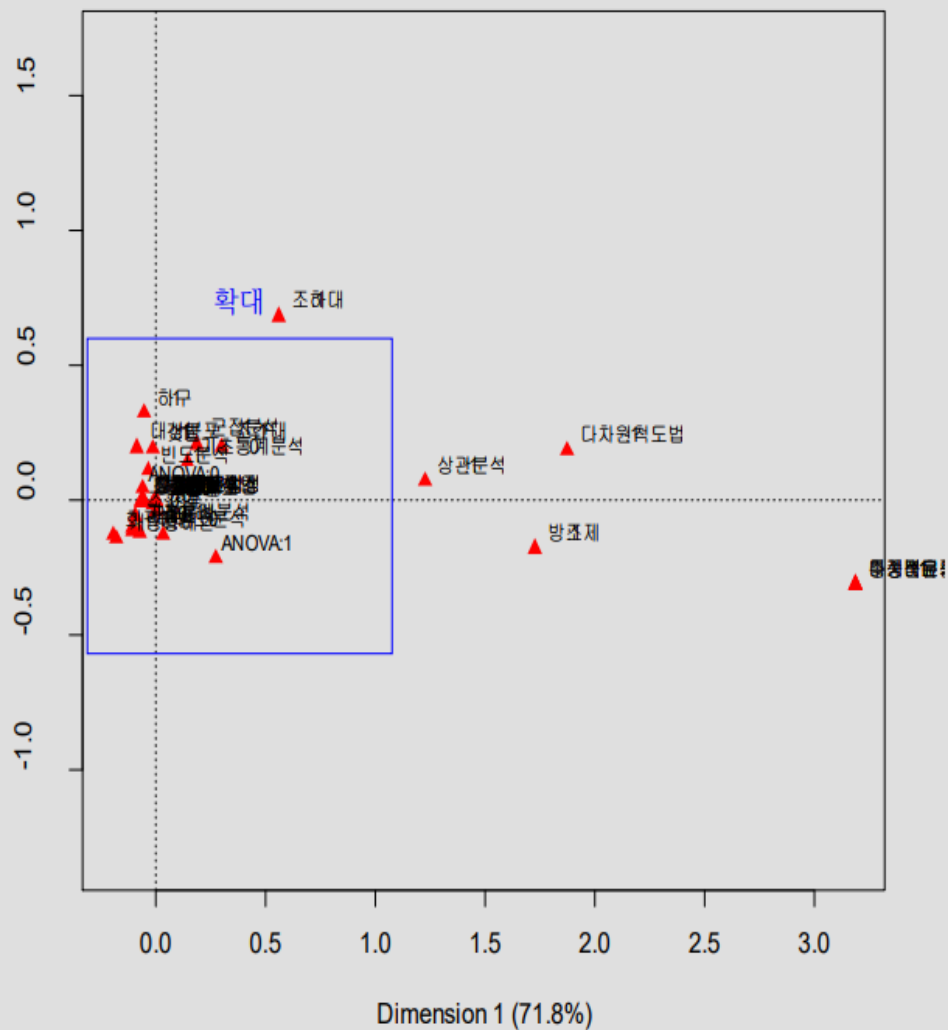
Seas

Habors

Key Words

Sea animals and plants





```
> apply(data[,c(1:11,45:51)],2,sum)
```

기초통계분석	빈도분석	군집분석	다차원척도법	주성분분석
35	32	21	2	1
ANOVA	통계적유의성	다중비교검정	상관분석	회귀분석
10	1	1	3	1
t검정	대상분포	조건대	조하대	방조제
1	2	14	1	2
갯벌	개방형해안	하구		
10	1	1		

통계기법과 Zonation의 연관성을 위한 다중 대응분석 그림

빈도 : apply(, sum)

Expansion

Relationship between Statistical Methods and Zonation Areas



- 1) 제1사분면 : 조간대 지역과 군집분석
- 2) 제2사분면 : 빈도분석과 갯벌 지역 연구
- 3) 제3사분면 : 조간대지역 외 : 기초통계분석, 회귀분석, t검정 등

확대 영역에 대한 다중 대응분석 그림 : library(ca), mja()

해 석

응용 3: JKSS 저널의 저자와 중심어의 SNA(library(sna))와 CRA(library(ca))

Choi et al.(2012) Social Network Analysis and It's Applications for Co-authors and Key-words Networks in the JKSS, Communications for Statistical Applications and Methods.

A140		Joong Jae Cho															
	Authors	BG	BH	BI	BJ	BK	Keywords		BN	BO	BP	BQ	BR	BS	BT	BU	
1		modified	ARCH	Beta proc	Markov ch	Second or	extreme vs	slope rot	compound	Jeffreys'	proportio	bootstra	Logistic	Markov ch	double	sampling	
2	Tae Sung Kim	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	Mi Hwa Ko	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	Hyo Il Park	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
5	Seung Hoe Choi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
6	Hae Kyung Kim	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
7	Young Lee	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
8	Jong Ho Bae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
9	Mee Seon Jeong	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
10	Choong Rak Kim	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
11	Byung Jin Choi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
12	Kee Young Kim	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
13	Jae Kwang Kim	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
14	Young Dai Kim	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
15	Jai Won Choi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
16	Balgobin Nandram	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
17	Narinder Kumar	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
18	H. J. Khamnei	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
19	Seung Chun Lee	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
20	Yoo Sung Park	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
21	Oe Sook Lee	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
22	Tae Yoon Kim	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
23	Jang Han Kim	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
24	Tai Sup Lee	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
25	Pen Hwang Liao	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
26	K. W. Kim	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
27	N. Y. Kim	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
28	Theodoros Nicolieris	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
29	Stephen G. Walker	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
30	Dal Ho Kim	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	
31	Sang Gil Kang	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	
32	Woo Dong Lee	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	

저자최종

저자가중최종

저자및핵심어최종

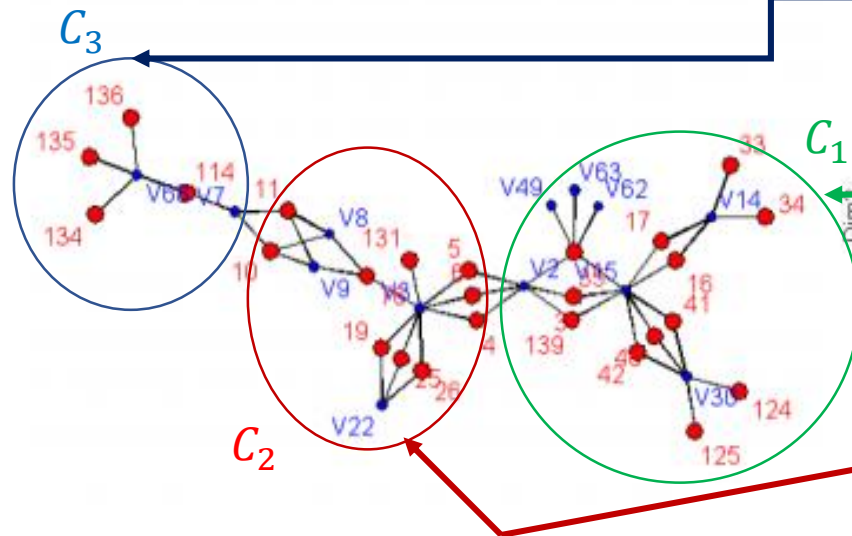
지역핵심어

Sheet1

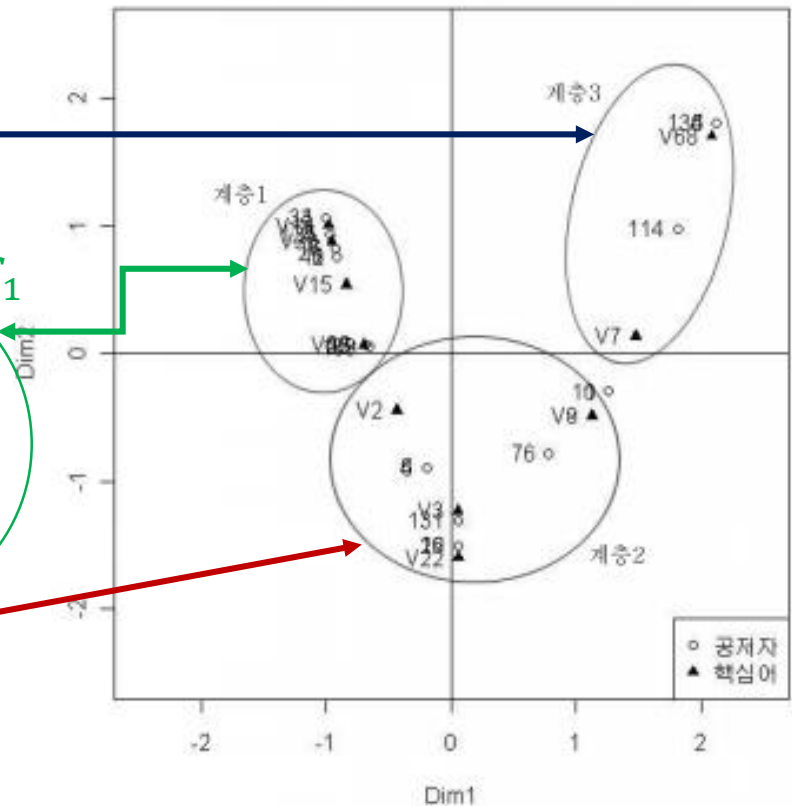
Sheet2

Sheet3

SNA Plot



CRA Map



Relationship between Authors and Keywords

계층	구분	저자 및 핵심어
C_1	저자	Narinder Kumar(16), Hyo Il Park(3), A. N. Gill(34), Seok Hoon Yun(35)
	핵심어	asymptotic normality(V15), Robust(V30), extreme value index(V49)
C_2	저자	Hae Kyung Kim(5), Young Lee(6), N. Y. Kim(26)
	핵심어	efficiency(V3), Multi-level rotation design(V8)
C_3	저자	Kee Young Kim(11), Yong Hee Lee(114), Jae Won Lee(136)
	핵심어	coefficient of variation(V7), likelihood ratio test(V8), bootstrap(V68)