

대학생 대상 정보 리터러시 연구의 키워드 네트워크 분석 및 토픽 모델링

소모임 1회차

Written by 60220311 박도현

When 2025.11.06.목

WHY?

AI 활성화로 인한 부작용 '할루시네이션'

최근 AI가 발전함에 따라 다양한 분야에서 AI를 사용하는 추세임
특히 대학 수업을 듣다보면 과제에 AI를 활용하는 경우가 많이 보임
그러나 AI가 주는 정보를 확인하지 않고 그대로 사용해
사실과 일치하지 않는 정보를 바탕으로 과제를 수행하는 학생들이 있음

이와 같은 모습을 통해 대학생의 리터러시 능력의 필요성을 느낌
따라서 현재 이와 관련해 어떤 연구가 진행되고 있는지 호기심이 생겼고,
해당 논문을 선정하게 됨.

ABSTRACT

정보 리터러시는 정보화 사회를 살아가는 모든 사람들에게 필요한 역량이지만, 대학생은 학업 수행과 직업 준비의 과정에서 정보 리터러시를 갖춰야 할 필요성이 특별히 높은 시기이다.

본 연구에서는 KCI에 등재된 대학생을 대상으로 진행한 정보 리터러시 관련 연구의 영문 초록을 대상으로 **빈도 분석**, **네트워크분석**, **토픽 모델링**을 수행하여 **대학생 대상 정보 리터러시 연구의 동향**을 파악하고자 하였다.

빈도분석과 키워드 네트워크를 분석하고 그 결과를 비교하여 주요 키워드와 후속 연구 주제를 도출하였으며, 토픽모델링 결과 8개의 하위 토픽을 도출하여 주 연구 분야를 관찰하였다. 대학생 대상 정보 리터러시는 주로 교육적 목적을 중심으로 연구가 진행되었으며, 간호 정보와 분석 모델 개발이 주요 하위 토픽으로 드러났다.

BACKGROUND KNOWLEDGE

LDA?

- 대량의 문헌을 비슷한 주제로 묶어낼 수 있는 특징을 찾는 토픽모델링의 한 모델
- 문서 집합에서 각 문서의 주제를 찾아내기 위한 통계적 모델링
- 문서 속 단어의 분포를 보고 문서가 어떤 내용을 다루고 있는지 추론
- LDA에서는 문서가 여러가지의 주제를 가질 수 있다고 가정함
 - = 한 문서가 어떤 주제들을 어떤 비율로 다루고 있는지 또한 추론함

즉 ! 컴퓨터가 긴 텍스트 속에서 자주 나오는 단어들을 기반으로
주제를 자동으로 찾아내는 도구로, 방대한 텍스트 속에서 어떤 주제가
중요한지 빠르게 파악 가능

BACKGROUND KNOWLEDGE

빈도분석?

- 키워드가 문서에서 출현한 빈도를 바탕으로 문서의 키워드 분포를 파악하는 기초적 분석 방법
- CF (Collection Frequency) : 빈도분석 지표 중 하나로, 문헌 집합에서 특정 단어가 나온 총 횟수
여기서는 연구 대상 논문들의 단어 각각의 총 횟수로 볼 수 있음
- CF는 특히 전체 데이터에서 자주 나오는 용어들의 중요성을 파악하는데 용이함

BACKGROUND KNOWLEDGE

네트워크 분석

- 개체 간의 관계를 파악하는 분석 기법,
여기서는 동시출현한 키워드들 간의 관계를 파악하는데 사용
- 연결 중심성 : 특정 키워드가 다른 키워드와 얼마나 연결되어있는지 정도
다른 키워드와 연결된 선이 많을 수록 높음
- 근접 중심성 : 특정 키워드와 다른 키워드간의 거리
- 매개 중심성 : 다른 키워드가 해당 키워드를 얼마나 경유하는지
한 키워드가 얼마나 서로 다른 그룹을 연결해주는 다리 역할을 하는지

INTRO

연구 배경

- 정보통신 기술 발전으로 인해 누구나 손쉽게 정보를 생산하고 공유할 수 있음
 - > 정보 생산량이 급격히 증가
- 이로 인해 정보의 편향, 가짜 정보, AI의 무분별한 정보 양산 등의 문제가 발생
- 이러한 오염된 정보의 위험에서 올바르게 정보를 습득하고 활용하려면
 - 정보의 사실 여부와 유용성을 판단할 수 있는 능력인 **정보 리터러시** 능력이 중요해짐
- 실제로 해외에서는 일찍이 관련 연구가 활발하게 이루어져 다양한 기준과 모델이 존재
- 특히 대학생 시기에 형성된 리터러시 능력은 향후 사회와 평생학습에 직결됨

INTRO

연구 목적

01. 우리나라에서도 리터러시 관련 연구는 꽤 진행됐으나,
분야가 다양해 동향이 모호함
= 이 연구에서는 "정보 리터러시" 로 한정지어 연구 동향을 파악할 필요성 느낌
02. KCI에 등재된 대학생을 대상으로 진행한 정보 리터러시 관련 연구의 동향을 파악
= 기존 연구 동향과 구별되는 특징을 파악하고자 함
03. 해당 주제에 대한 주요 키워드를 파악하고, 후속 연구 주제 제언을 목표로 함

PRECEDING RESEARCH

리터러시의 개념 변화

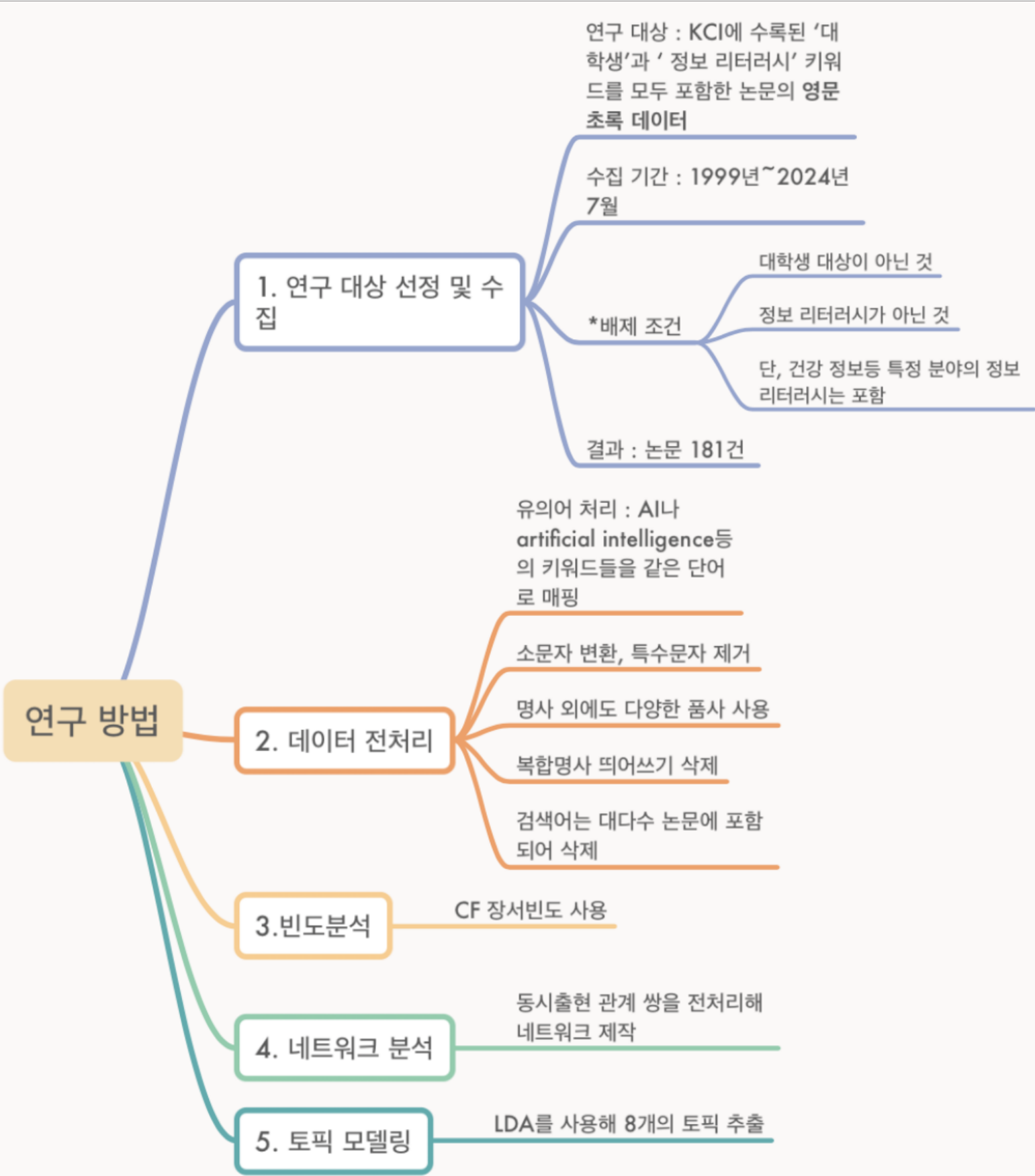
초기 : '문맹'의 반대 의미로 읽고 쓸 수 있는 능력을 의미

현재 : 특정 주제, 혹은 미디어를 '읽을' 수 있는 능력, 특정 분야에 대한 지식이나 경쟁력으로 의미 확대

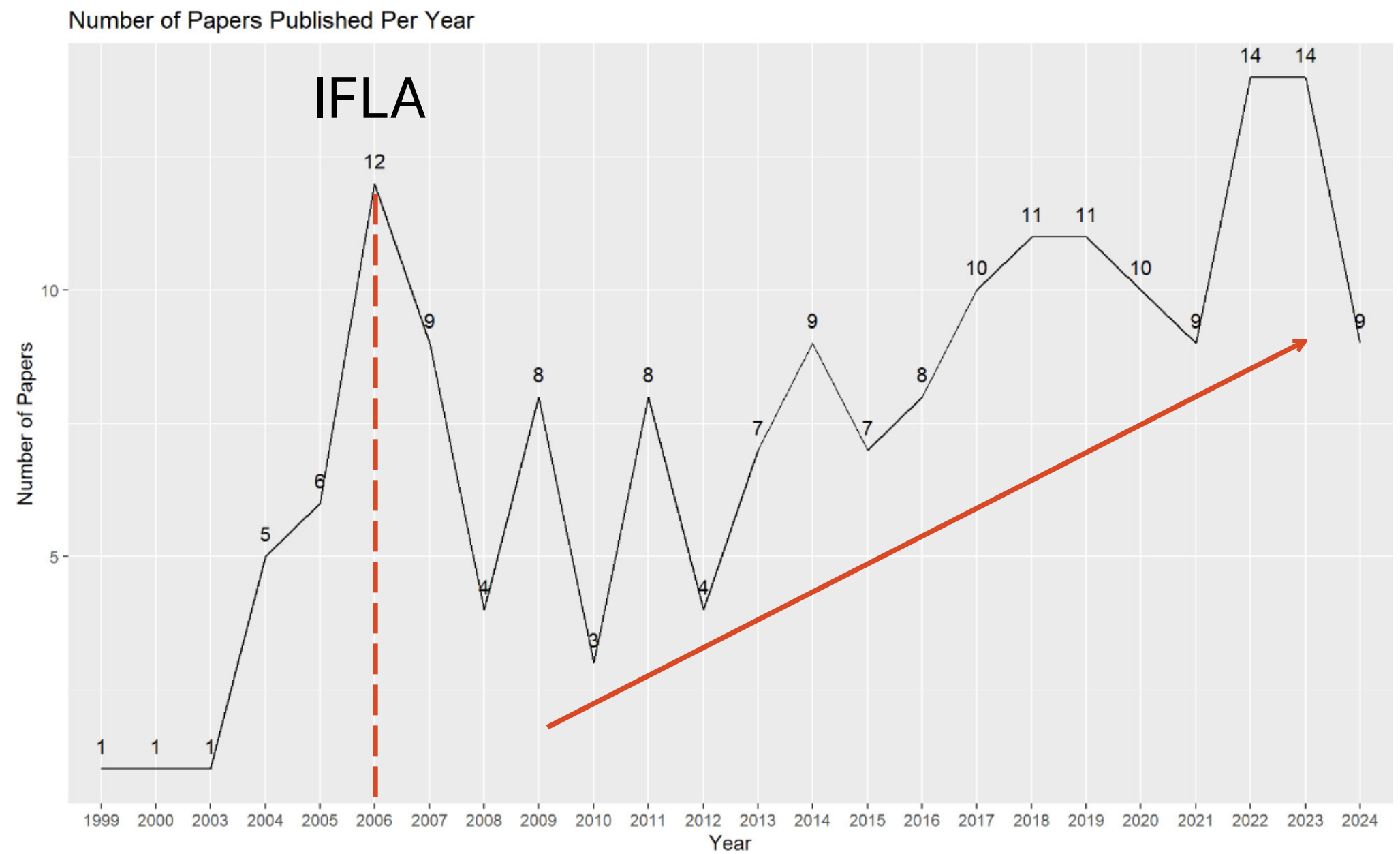
리터러시의 현황

- 정보 기술의 발전으로 인해 인터넷을 비롯한 디지털 환경에서 리터러시 개념이 중요해지는 추세
- 독서, 정보, 컴퓨터, IT 등 특정 주제와 리터러시가 결합된 복합 리터러시가 사용되고 있음

METHODOLOGY



분석대상 논문 발행 현황



〈그림 1〉 대학생 대상 정보 활용능력 연구 논문의 연도별 발행 현황

키워드 빈도 분석

- 노란색 : 교육 및 학업과 관련된 키워드가 가장 높은 빈도로 나타남
= 리터러시 교수, 학습 관련 연구가 많기 때문
- 초록색 : 학생의 역량과 관련된 어휘
= 학업 수행 능력의 일환으로 정보 리터러시 파악,
+ 리터러시와 학업 수행 능력의 연관성을 연구한 논문들
- 핑크색 : 간호 및 보건 의료 관련 키워드
= 간호 및 보건의료 분야의 정보리터러시 연구가 유의미하게 이뤄짐
- 보라색 : 디지털 환경과 관련된 키워드
= 코로나 이후 온라인에서 대면 전환하는 시기에
정보 리터러시 대두

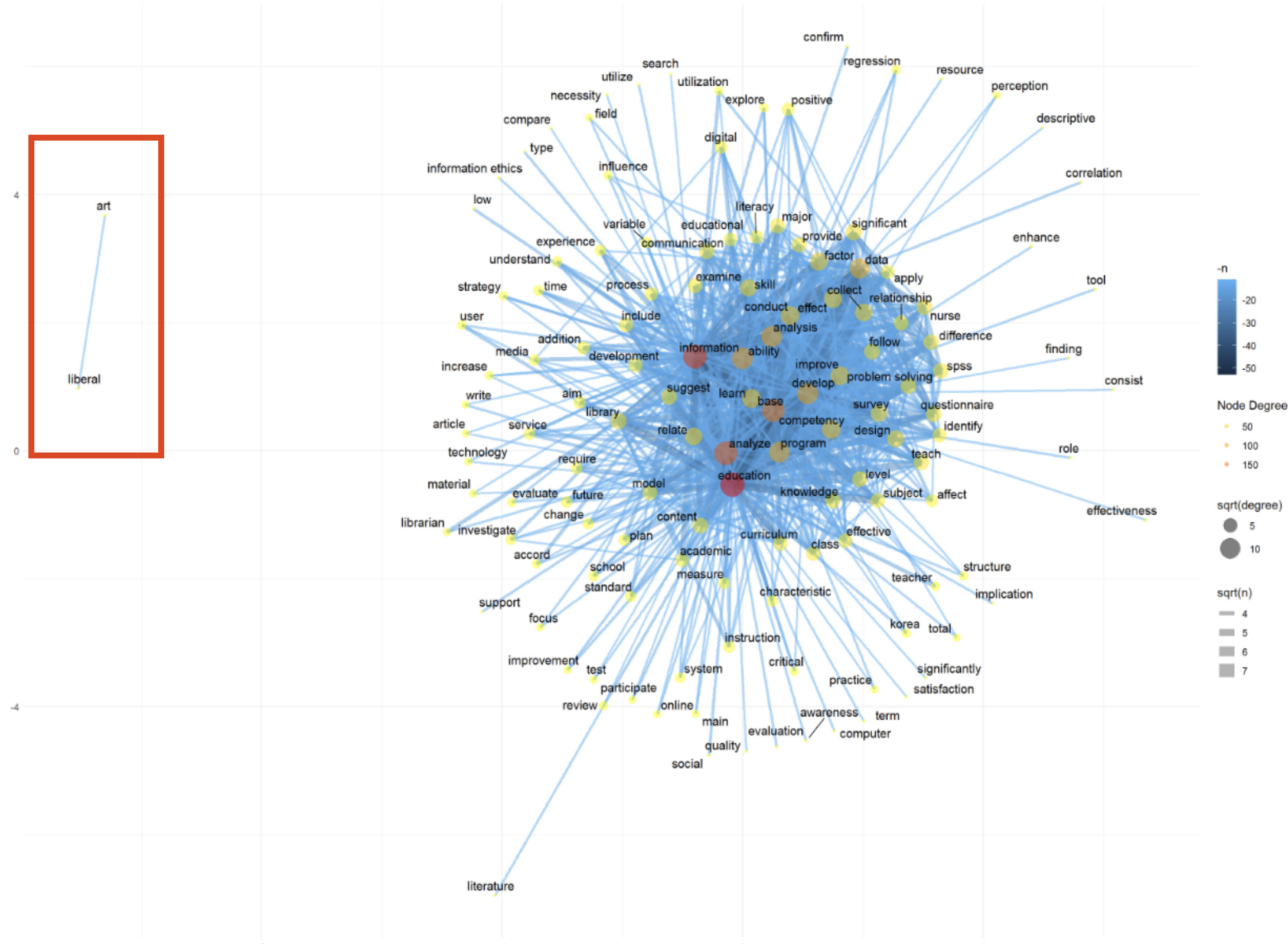
〈표 2〉 빈도 분석 결과

키워드	빈도	키워드	빈도	키워드	빈도	키워드	빈도
education	312	problem solving	73	instruction	44	influence	34
information	299	communication	72	media	44	information literacy instruction	34
competency	229	service	69	online	44	measure	34
library	149	improve	68	collect	43	school	34
learn	146	survey	66	include	43	experience	33
ability	142	content	65	librarian	43	require	33
nurse	134	level	65	questionnaire	42	characteristic	32
literacy	119	knowledge	64	examine	41	learner	32
skill	118	conduct	60	korea	41	satisfaction	32
base	116	relate	59	nursing information literacy	41	support	31
program	115	major	57	critical	40	utilization	31
analyze	101	class	55	item	40	internet	30
analysis	97	subject	55	perception	40	material	30
teach	96	process	53	relationship	40	plan	30
effect	94	development	52	variable	40	read	30
model	94	suggest	52	standard	39	source	30
develop	89	provide	50	change	38	awareness	29
curriculum	84	difference	49	core	38	direct	29
factor	83	significant	47	job	38	effective	29
teacher	83	test	47	positive	37	evaluate	29
academic	82	user	47	computer	36	strategy	29
digital	82	follow	46	identify	36	technology	29
design	79	system	46	korean	36	affect	28
write	75	apply	44	understand	35	art	28
data	74	educational	44	future	34	concept	28

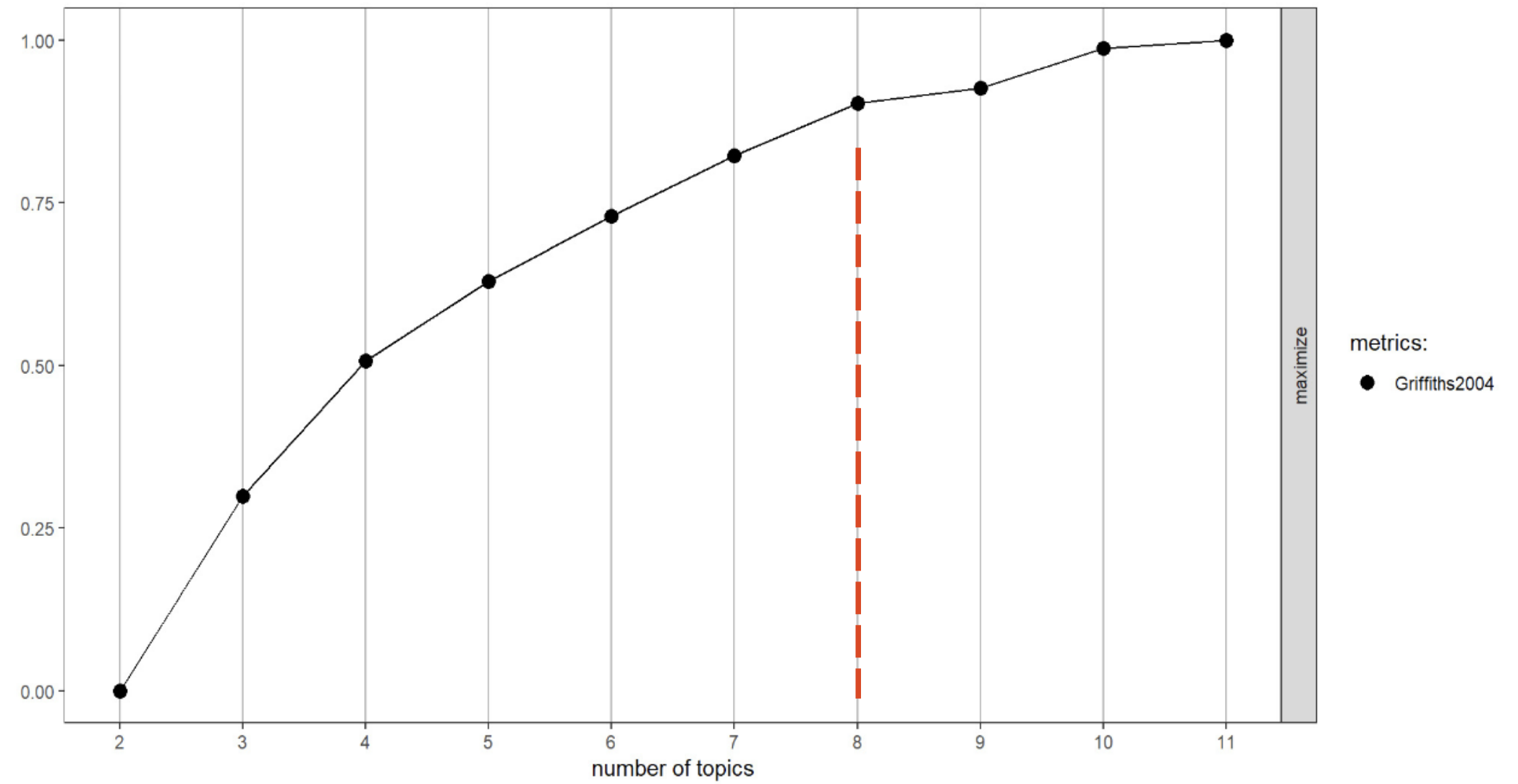
네트워크 분석

- 키워드 빈도가 가장 높은 Education이 연결, 근접, 매개 중심성이 가장 높았음 이 외에도 information, analyze, base등이 주요 키워드
- art 와 liberal 키워드의 경우 근접중심성은 높는데 매개 중심성과 연결 중심성이 매우 낮음 -> 네트워크에서 중요한 역할 안함.
- 60%의 키워드가 엣지가 하나임.
= 40%의 소수 키워드에 중심성이 집중된 구조임
- 인문 사회학 분야의 정보 리터러시 연구미비
문헌 리뷰나 기술적 연구등의 방법론이 상대적으로 잘 안쓰임
- 인문학 글쓰기에 있어 정보 리터러시의 역할, 온라인 정보 활용에서의 정보 윤리와 리터러시 등의 주제 도출 가능

Keyword Co-occurrence Network



토픽 모델링



〈그림 3〉 Griffiths2004 활용 토픽 수 산출

토픽 모델링

1. 대학생 정보 리터러시 교육 및 모델 설계

2.대학의 정보 리터러시 교육

3.온라인 정보 분석 및 평가

4.디지털 및 미디어 리터러시와 글쓰기

5.학습 능력 및 문제 해결을 위한 정보 리터러시

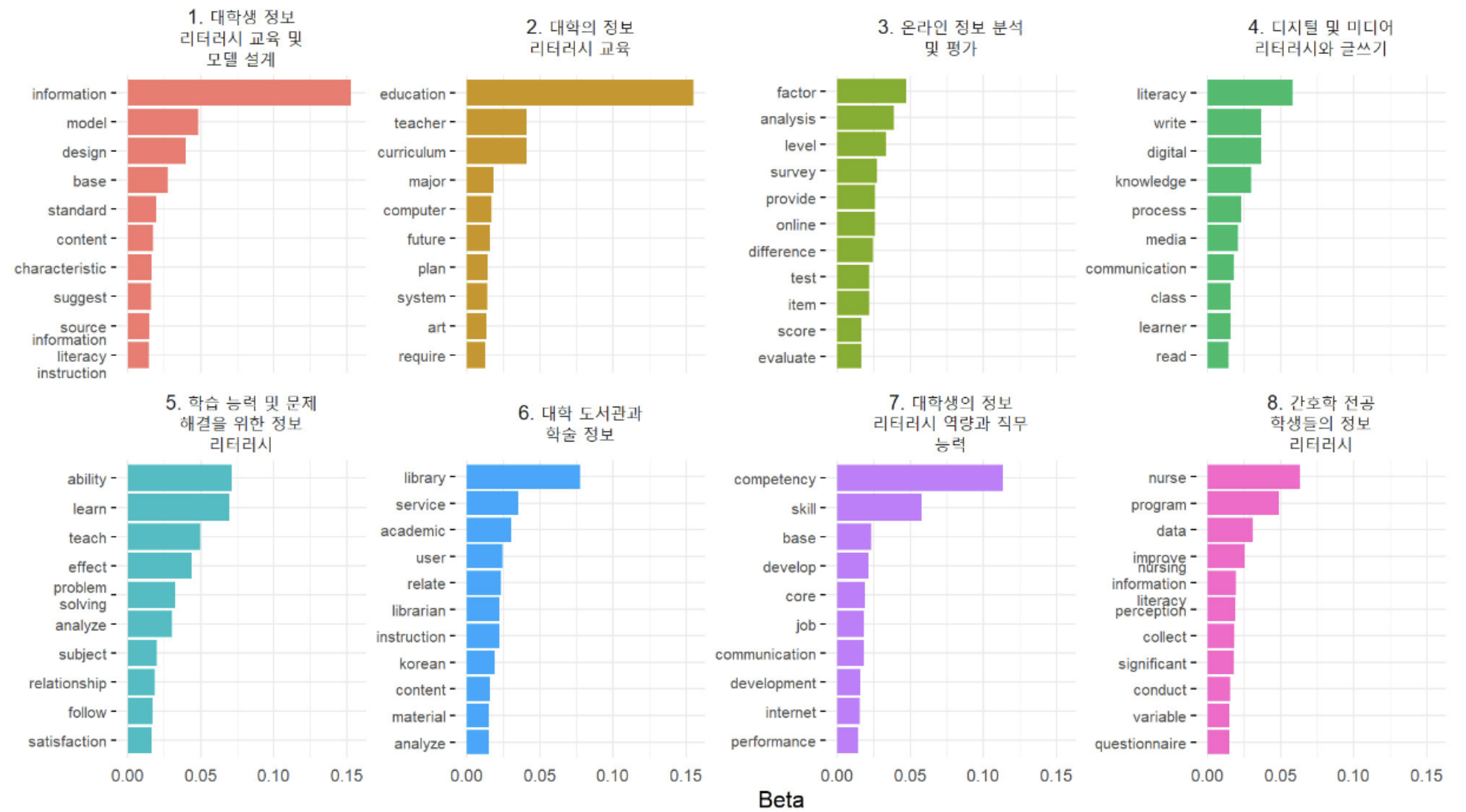
6.대학 도서관과 학술 정보

7.대학생의 정보 리터러시역량과 직무 능력

8.간호 및 건강 정보 리터러시

= 8개의 토픽이 비슷한 확률로 분포

Top 10 Keyword for Each Topics



CONCLUSION

연구 결과

8개의 토픽이 도출되었으며, 현재까지는 '대학 도서관과 학술 정보' (17.1%)와 '교육 및 모델 설계' 관련 연구가 가장 활발함

간호학을 포함한 특정 전공 분야의 정보 리터러시 연구는 다른 주제와 독립적인 토픽으로 나타날 만큼 활발하게 진행되고 있음.

학습 능력 및 문제 해결' (8.8%) 및 '직무 능력' (8.8%) 관련 연구는 향후 후속 연구가 활발히 이루어질 수 있는 분야임

CONCLUSION

의의 및 한계

특정 분야에 한정하지 않고 대학생의 정보 리터러시 관련 연구를 학제적으로 검토했다는 점에서 의의

디지털, '미디어' 등 특정 분야 리터러시를 정보 리터러시와 구분하는 논문을 분석 대상에서 의도적으로 제외하여 전반적인 리터러시 연구 동향이 명확하게 드러나지 않음

시계열 연구를 진행하지 않아 연구 동향의 변화를 명확하게 살피지 못했다는 한계가 있음

**THANK YOU
FOR LISTENING**