

Working Paper 2019-01

빅데이터 기반의 직업능력개발 연구동향 분석과 시사점

오호영·정지운·홍광표·이은혜·조영민

2019. 4.

 **한국직업능력개발원**

Korea Research Institute for
Vocational Education & Training

빅데이터 기반의 직업능력개발 연구동향 분석과 시사점

오호영, 정지운, 홍광표, 이은혜, 조영민¹⁾

< 목 차 >

I. 연구 배경 및 목적	2
II. 연구 방법	3
III. 분석 결과	7
IV. 결론 및 시사점	19
참고문헌	23

연구의 개요

- 직능원에서는 직능원이 개원 이래 20년간 수행하여온 연구보고서의 주제와 핵심단어를 빅데이터 분석 방법을 활용하여 심층적으로 분석하면서 직능원과 밀접한 관련성을 갖고 있는 직업능력개발 분야의 대표적인 학술지와의 연구동향에 대한 비교분석을 통해 향후 연구주제의 발굴과 연구방향성의 정립에 있어서 중요한 기초자료로 활용하고자 하였음. 분석결과 직능원의 기본연구의 연구동향과 유관학술지들의 연구동향은 뚜렷한 차이점을 보이고 있으며 이는 분석대상 학술지의 특성, 정책연구기관으로서 직능원의 사회적 책무성과 연구지향에서의 차이를 반영하는 것으로 해석할 수 있음.

- 주제어: 빅데이터(big data), 토픽모델링, LDA분석, 직업능력개발

1) 한국직업능력개발원 오호영·정지운·홍광표·이은혜, 연세대학교 상경대학 경제학과 조영민

I. 연구 배경 및 목적

4차 산업혁명, 인구구조 변화 등의 급격한 사회경제적 환경변화가 진행되는 가운데 우리 사회가 어떻게 급변하는 새로운 환경에 적합한 제도, 정책, 인프라 등을 체계화하고 혁신해 나갈 것인가는 사회과학 분야 연구자들의 주요한 관심사가 되고 있다. 산업화, 개발연대에 기틀을 마련한 우리나라의 사회시스템은 소득 3만불의 선진화 시대로 접어들면서 근본적 변화요구에 직면하는 한편으로, 세계적으로 유례를 찾기 힘든 고도의 저출산과 고령화로 인한 급격한 인구구조의 변화, 인공지능, 빅데이터, 로봇 등의 분야에서 이루어지는 눈부신 기술혁신에 따른 노동과 숙련의 급격한 변동은 사회시스템의 혁신을 더 이상 미루기 어려운 과제로 만들고 있다.

사회를 탐구대상으로 삼는 사회과학 연구의 특성상 사회현상의 급격한 변화는 관련 연구에 대한 연구자들의 관심을 촉발하고 정책적 대응방안 마련과 학술적 연구방법론에 대한 연구를 활성화하는 계기로 작용하게 된다. 예컨대, 동일한 고등교육 분야에 대한 연구라 하더라도 한국과 같이 대졸 청년층의 취업난이 심각한 상황에서는 대졸청년층의 취업난의 원인을 진단하고 일자리 창출, 산업계 수요 관점에서의 대학 인력양성 등과 같은 실천적인 연구에 대한 사회적 수요가 높아지고 연구자들은 논문과 보고서 등을 통해 이에 반응하게 된다. 한편, 미국과 같이 대학진학률이 상대적으로 낮은 국가에서는 대학진학의 경제적 효과분석, 대학생에 대한 재정지원 시스템 개혁, 대학중퇴율을 낮추기 위한 방안연구 등이 활발하게 이루어지는 경향이 있다.

한국직업능력개발원(이하 직능원)은 “직업교육훈련정책 및 자격제도에 관한 연구와 직업교육훈련 프로그램의 개발·보급 등 직업능력개발에 관한 업무를 효율적으로 수행함으로써 직업교육훈련의 활성화 및 국민의 직업능력 향상에 기여함을 목적으로” 1997년 설립되었다.(한국직업능력개발원 정관 제2조) 1997년은 외환위기로 인한 국가적 위기감이 높았던 시기로서 실업문제가 사회적 이슈가 되었으며 직능원은 이에 효과적으로 대응하기 위하여 직업능력개발 정책을 선도하고 효율화하기 위한 각종의 연구사업을 수행하여 왔다. 직능원은 시대적 요구에 부응하여 국가인적자원개발, 진로교육, 자격, 고졸취업, NCS 학습모듈 등 아젠다 설정과 연구영역의 확장을 통해 학술적, 정책적으로 기여하여 왔다.

매년 다양한 경로의 외부 의견수렴, 정부의 국정과제 및 부처의 요구, 원내 연구진의 연구의견 등을 반영하여 과제발굴과 연구사업을 적극적으로 수행하여 왔으나, 직능원 외부 연구동향에 대한 체계적 분석과 이에 기초한 연구방향 재정립 등은 상대적으로 부족하였던 것이 사실이다. 이는 직능원에 국한된 문제라기보다는 연구동향을 체계적으로 분석하기 위한 방법론, 인프라 등이 제대로 못 갖춰진 상태에서 외부 전문가들의 의견을 통해 대처할 수밖에 없었던 국책 연구기관의 공통된 한계로 생각된다.

최근 빅데이터(big data) 분석기법의 도입으로 인하여 정형화된 수치 데이터 분석에 국한되어 왔던 통계분석이 대규모의 비정형 데이터로 확장됨으로써 텍스트자료, 영상자료, 음성자료 등도 가치 있는 데이터로 분석의 외연이 확장되고 있다. 이에 직능원이 개원 이래 20년간 수행하여온 연구보고서와 직능원과 밀접한 관련성을 갖고 있는 직업능력개발 분야의 대표적인 학술지와의 연구동향을 비교분석하고 향후 연구주제의 발굴과 연구방향성의 정립에 있어서 중요한 기초자료로 활용하고자 한다. 직능원의 연구는 국책연구기관으로서 사회적 책무를 다하기 위해 정책수요를 적극적으로 반영해야 함과 아울러 새롭게 제기되거나 제기될 사회이슈들을 선도적으로 발굴하고 정책적, 학술적으로 사회적 논의를 주도해나갈 필요성이 있으며, 본 연구를 통하여 외부 전문가들의 연구동향과 직능원의 연구동향을 비교분석하고 향후 직능원의 연구방향 좌표를 설정하는데 기여하고자 한다.

II. 연구 방법

1. 분석 자료

이 연구는 직능원의 연구 동향 및 교육훈련·고용 분야의 국내 연구 동향을 파악하기 위해 직능원 개원이래 수행된 기본연구보고서 영문초록과 교육훈련·고용 분야의 국내 주요 학술지 6종에 게재된 논문의 영문초록을 분석 자료로 활용하였다. 시계열분석을 위해 한국직업능력개발원의 1998~2018년도까지 발간된 기본보고서와 1990~2018년 기간 중에 발간된 국내 학술지의 논문 초록을 수집하였으며, 데이터 전처리 과정이 복잡한 국문 초록 대신 영문 초록을 분석 대상으로 하였다. 학술지는 한국학술정보(KISS) DB에서 최근 1년간 직능원 연구

자들의 이용 횟수가 높은 상위 5개의 국내 학술지 중 직능원의 주요 연구 분야와 관련성이 높은 학술지 4종 『교육공학연구』, 『직업교육연구』, 『역량과 성과관리연구』, 『기업교육과 인재연구』를 선정하였다. 그리고 고용/노동시장 연구 분야를 보강하기 위해 『한국노동경제논집』을 추가하였으며, 직능원에서 발간하는 학술지 『직업능력개발연구』²⁾를 더하여 총 6개의 학술지를 분석 대상으로 선정하였다.

수집한 영문 초록은 총 3,061개로 직능원의 기본연구보고서 영문 초록 803개, 『교육공학연구』는 734개, 『직업교육연구』는 523개, 『역량과 성과관리연구』는 336개, 『직업능력개발연구』는 238개, 『기업교육과 인재연구』는 233개, 『한국노동경제논집』은 194개이다. 시기별로는 1999년 이전에 발간된 연구물의 영문 초록이 236개, 2000~2005년은 574개, 2006~2010년에는 806개, 2011~2015년은 960개, 2016년~2018년은 485개이다. 참고로 『직업교육연구』, 『역량과 성과관리연구』, 『한국노동경제논집』은 2000년 이후에 창간되어 창간호부터 2018년까지 발간된 논문 초록을 수집하였다.

<표 1> 분석 자료

(단위: 개)

	직능원 기본보고서	교육공학 연구	직업교육 연구	역량과 성과관리 연구	직업능력 개발연구	기업교육과 인재연구	한국노동 경제논집	합계
1999년 이전	64	145	-	-	18	9	-	236
2000~2005년	225	177	-	-	67	45	60	574
2006~2010년	216	161	148	119	126	18	18	806
2011~2015년	209	158	265	150	13	94	71	960
2016년 이후	89	93	110	67	14	67	45	485
합계	803	734	523	336	238	233	194	3,061

2. 분석 방법

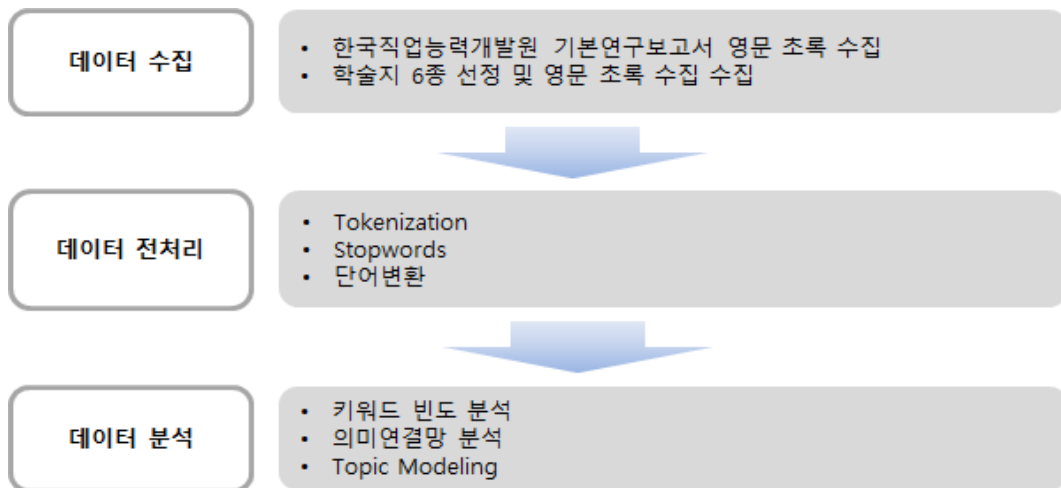
연구 절차는 [그림 1]과 같이 데이터 수집, 데이터 전처리와 분석 순으로 이루어졌다.

데이터 수집 단계에서는 직능원의 기본연구보고서와 학술지 『직업능력개발연

2) 『직업능력개발연구』는 중간에 『고용직업능력개발연구』로 학술지명이 변경되었음. 이 연구에서는 『고용직업능력개발연구』에 실린 논문도 『직업능력개발연구』에 포함시켜 분석함.

구』의 경우 내부 DB에서 직접 영문 초록을 추출하였으며, 나머지 5종의 학술지는 한국학술정보(KISS) 논문검색서비스 홈페이지에 게시된 영문 초록을 수집하였다. 분석에는 R 프로그램을 사용하였으며, 데이터 전처리 단계에는 텍스트마이닝에서 보편적으로 활용되는 ‘tm 패키지’, 토픽모델링에는 ‘topicmodels 패키지’가 적용되었다. 본격적인 분석에 앞서 데이터 전처리 단계에서 수집한 문서 자료를 분석에 활용할 수 있도록 단어수준으로 쪼개는 토큰화(Tokenization) 과정을 거쳤으며, 분석에 불필요한 단어를 제거하는 Stopwords 과정을 거쳐서 텍스트를 정제하였다. 그리고 의미는 같으나 다른 형태로 표현된 단어들을 하나의 공통 단어로 변환하여 데이터를 분석에 활용할 수 있도록 처리하였다.

[그림 1] 연구절차

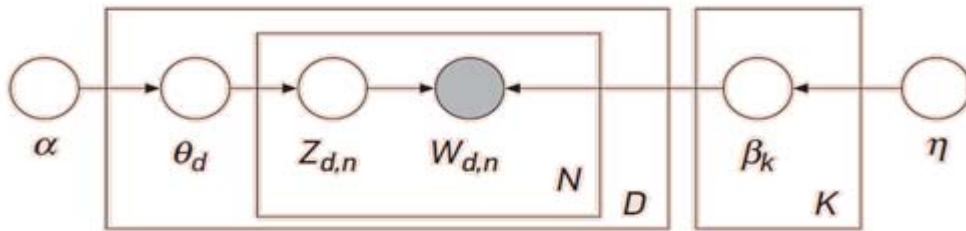


데이터 정제 후에는 데이터의 출현 빈도를 분석하여 키워드를 추출하는 키워드 분석을 실시하였고, 주요 키워드 간의 관계를 파악하기 위하여 의미연결망 분석을 실시하였다. 의미연결망 분석에서 연속으로 등장하는 두 단어인 바이그램(bigrams)을 기준으로 분석하였으며, 노드는 키워드이며, 노드와 노드를 연결하는 링크는 한 문서(초록)에 동시 출현했는지의 여부에 따라 발생하도록 설정하였다. 시각적으로 의미연결망 분석이 가능하도록 직관적인 해석을 위해 바이그램 빈도수를 기준으로 100개 이상 등장하는 바이그램을 기준으로 의미연결망 분석을 하였다.

방대한 양의 연구물이 어떤 주제로 분류될 수 있는지를 살펴보기 위해 토픽

모델링(topic modeling)의 LDA(Latent Dirichlet Allocation) 방법을 사용하였다. 토픽 모델링은 실제 관찰 가능한 문서 내 단어를 사용해서 토픽의 단어 분포와 문서의 토픽 분포를 추정하는 과정이며, LDA는 문서-토픽, 단어-토픽으로 문서를 재현할 때 얼마나 원 문서와 일치하는지를 반복적으로 수행하면서 원 문서에 가장 근접한 문서-토픽, 단어-토픽 관계를 도출하는 기법이다. LDA 그래픽 모델링은 [그림 2]와 같다. 하이퍼파라미터인 α 와 η 가 주어지면 α 를 통해 θ 를 구하고 θ 로 각 단어의 토픽을 나타내는 Z 을 생성한다. 그리고 η 를 통해 β 를 구하면 Z 값과 β 값에 따라 단어 W 가 결정되는 알고리즘으로 되어 있다. 이 연구에서는 토픽 모델링 LDA를 사용하여 도출한 주제로 총 3,061개 영문 초록을 분류하고, 출처별, 시계열별 주요 연구 주제의 변화를 비교하여 연구 동향을 분석하였다.

[그림 2] 토픽 모델링 LDA 그래픽 모델링



- K: 토픽 수
- N: (특정) 문서의 단어 수
- D: 문서 수
- α : 문서별 토픽 K의 Dirichlet prior weight, θ 값을 결정하는 파라미터
- θ_d : 문서별 토픽의 비율
- β_k : 토픽별 단어 w의 생성 확률
- $Z_{d,n}$: 문서 d의 n번째 단어의 토픽
- $W_{d,n}$: 문서 d의 n번째 단어

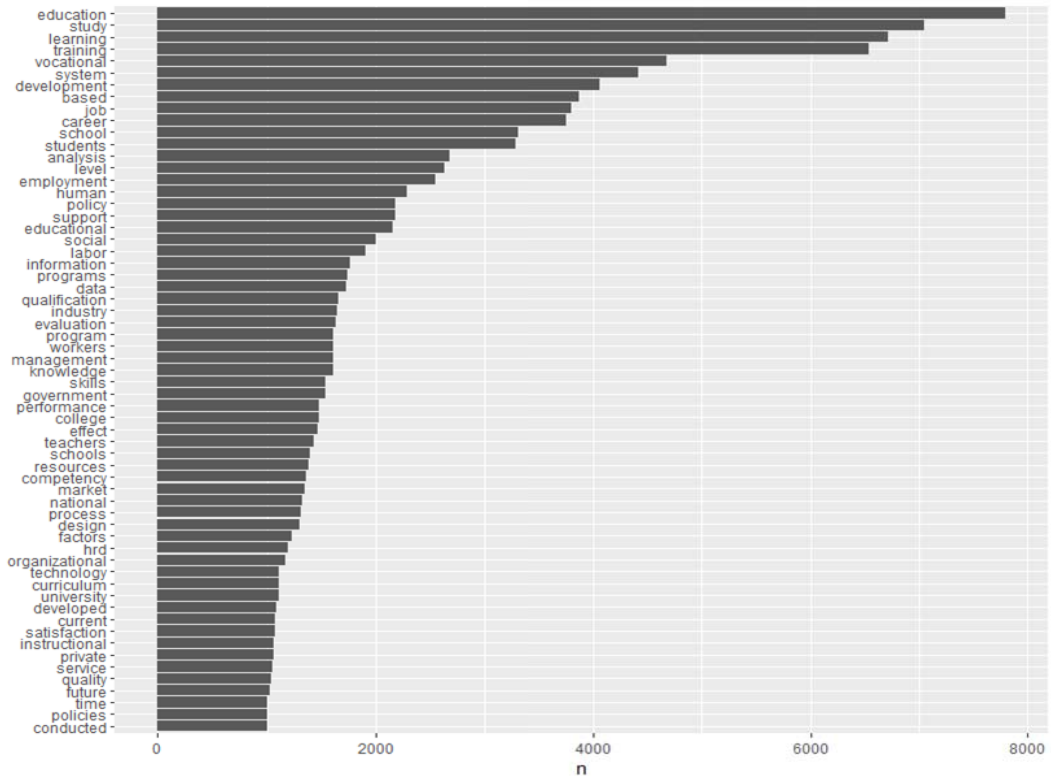
자료: Blei, D. M.(2012). Probabilistic Topic Models, *Communications of the ACM*, 55(4), pp.81, Figure 4. The graphical model for latent Dirichlet allocation.; 김상겸(2016), 『토픽모델링을 이용한 국내 산업공학 연구동향 분석』, 서울과학기술대학교 석사학위논문.

Ⅲ. 분석 결과

1. 키워드 분석 결과

키워드 분석을 위해 3,061편의 영문 초록에서 빈도가 1,000번 이상 나타난 단어를 추출하였다. 출현 빈도가 가장 높은 키워드는 ‘education’이며, 그 다음은 ‘study’, ‘learning’, ‘training’ 순으로 나타나 교육과 관련된 내용이 가장 많이 다루어지고 있음을 알 수 있다. 그 다음 출현 빈도가 4,000~5,000번 사이에 있는 단어는 ‘vocational’, ‘system’, ‘development’로 직업교육과 제도/체계, 개발과 관련된 주제 또한 중요하게 다루어지고 있는 것으로 나타났다. 3,000~4,000번의 출현 빈도를 보이는 단어는 ‘based’, ‘job’, ‘career’, ‘school’, ‘student’ 등으로 학교와 학생을 대상으로 하는 주제들이 자주 등장하였으며, 진로/직업 분야 또한 자주 연구되는 분야임을 알 수 있다. 2,000~3,000번의 출현 빈도를 보이는 단어로는 ‘employment’, ‘human’ 등이 있으며 고용 및 인적자원에 대한 연구주제를 반영한 것으로 보인다. 1,000~2,000번의 출현빈도를 보이는 단어들이 다수 나타나고 있으며, ‘labor’, ‘information’, ‘qualification’, ‘future’, ‘time’ 등으로 이는 노동시장, 정보기술 등 직업능력개발을 둘러싼 환경에서 진행되고 있는 변화, 특성 혹은 미래전망 등에 관한 연구관심을 나타내주며, 자격이 직업능력개발과 노동시장을 연결하는 핵심매개로서 중요성이 강조되고 있는 것으로 해석이 가능하다.

[그림 3] 교육훈련·고용 연구 분야 주요 키워드



[그림 4]는 교육훈련·고용 연구 분야 출현빈도가 높은 단어를 중심으로 워드 클라우드를 작성한 것이다.

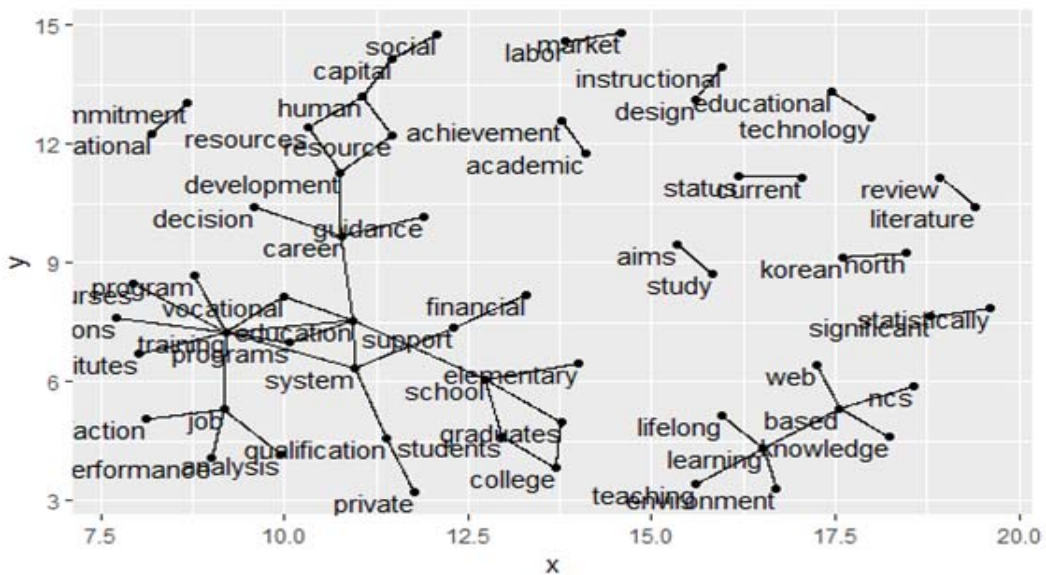
[그림 4] 교육훈련·고용 연구 분야 워드 클라우드



[그림 5]에는 교육훈련·고용 연구 분야 키워드 간 의미연결망이 제시되어 있는데, 주요 키워드로 구성된 의미연결망을 보면 ‘education’ 을 중심으로 연결된 집단1과 ‘learning’, ‘based’ 를 중심으로 연결된 집단2, 그리고 나머지 주변부로 나눌 수가 있다. ‘education’ 단어 상단의 연결망은 진로지도/인적자원개발/사회적자본과 연결되어 있으며, 왼쪽은 직업교육훈련 프로그램과 자격과 연결되어 있다. ‘training’, ‘education’ 은 공통적으로 ‘vocational’, ‘system’ 과 연결되어 있으며, 상단에서는 ‘career guidance’, 하단에서는 ‘qualification’ 과 연결되어 있다. 직업능력개발을 ‘vocational education & training’ 이라고 본다면 이러한 능력개발은 진로지도와 자격을 매개로하여 노동시장과 연결되고 개인의 인적자원개발을 촉진하는 것으로 해석이 가능하다.

하단 오른쪽 연결망은 학습, 능력개발의 구체적인 방식과 관련된 키워드로 구성되어 있다. ‘learning’, ‘based’ 로 연결된 집단2의 상단은 ‘web’, ‘ncs’, ‘based’ 로 연결되어 있으며, 하단은 ‘lifelong’, ‘learning’, ‘teaching’, ‘environment’ 로 연결되어 있다. NCS기반의 직업능력개발체제가 직업교육과 직업훈련에서 자리잡아 감에 따라 이에 대한 연구가 활발히 추진되고 있으며 인터넷 기반의 학습이 평생교육의 새로운 양상으로 전개됨을 보여주고 있다.

[그림 5] 교육훈련·고용 연구 분야 키워드 간 의미연결망



2. 토픽 모델링 결과

직능원의 기본과제 보고서 요약과 관련 연구 분야의 주요 학술지 초록을 대상으로 토픽모델링을 수행하여 20개의 주제와 각 주제별 상위 단어를 20개씩 추출하였다. LDA기법을 적용하여 토픽을 추출하였는데, LDA 기법은 문서(즉, 각각의 보고서 혹은 논문 요약문)는 토픽들의 결합으로 생성되며, 다시 이러한 토픽이 각각의 토픽확률분포에 의거하여 문서에서 단어를 생성하는 것으로 본다. 이에 따르면 어떤 문서가 주어졌을 때, 우리는 해당 문서의 단어분포를 생성하는 최적의 토픽확률분포를 역으로 계산해낼 수 있게 된다. 각 토픽별 상위 20개 단어는 각각의 추출된 토픽이 어떤 특성을 갖는지 내용정보를 제공해주게 된다. 연구진에서는 각 토픽의 상위단어를 중심으로 하위의 단어들과 연계하여 연구자가 이해할 수 있는 명칭을 각각의 토픽에 부여하는 작업을 수행하였다.

그 결과, 직능원의 기본과제와 교육훈련·고용 분야 학술지는 총 20개의 토픽으로 구성되어 있는 것으로 나타났다. Topic1-학교 교육, Topic2-직업교육 제도 및 정책, Topic3-지식·역량, Topic4-일자리·고용, Topic5-심리적·정서적 유인, Topic6-조직관리, Topic7-자격, Topic8-교사, Topic9-학생, Topic10-인력 수요·공급, Topic11-사회정책과 청년고용, Topic12-직업훈련 프로그램, Topic13-인적자원개발 정책, Topic14-개발 협력, Topic15-숙련과 노동시장, Topic16-프로그램 평가, Topic17-역량과 성과관리, Topic18-온라인·평생교육, Topic19-교육공학, Topic20-진로지도 등이다. 이러한 토픽의 명칭은 토픽별 단어빈도를 기반으로 연구진이 주관적으로 부여한 것이기 때문에 객관성과 합리성을 확보하기 위해서는 이 보고서를 출발점으로 좀 더 포괄적인 논의를 바탕으로 전문가의견을 수렴하는 과정이 필요해 보인다.

빅데이터 기반의 직업능력개발 연구동향 분석과 시사점 (오호영 외)

<표 6> 교육훈련·고용 연구 분야 주제

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
	학교 교육	직업교육 제도 및 정책	지식·역량	일자리·고용	심리적· 정서적 유인	조직관리	자격	교사	학생	인력 수요·공급
1	school	education	knowledge	job	factors	organizational	system	teachers	group	demand
2	high	vocational	process	employment	satisfaction	effect	qualification	teaching	students	industry
3	students	system	problem	work	efficacy	relationship	private	teacher	groups	manpower
4	college	society	activities	jobs	significant	employees	national	subject	task	growth
5	schools	educational	problems	occupations	effect	leadership	qualifications	class	types	service
6	university	higher	based	working	motivation	positive	management	school	significant	future
7	curriculum	policy	practice	occupational	level	commitment	based	contents	control	industries
8	graduates	term	stage	women	achievement	organization	quality	method	cognitive	supply
9	vocational	current	solving	occupation	relationship	team	operation	studies	test	rate
10	colleges	long	sharing	people	influence	performance	standards	students	type	economy
11	educational	institutions	concept	index	perceived	behavior	certification	methods	difference	higher
12	level	industrial	individual	status	network	goal	government	elementary	collaborative	employment
13	specialized	based	understanding	female	academic	mediating	professional	application	interaction	change
14	student	streng	community	employed	social	emotional	systems	textbook	achievement	economic
15	universities	programs	business	time	perception	implications	recognition	field	discussion	workforce
16	year	order	members	foreign	attitude	psychological	current	subjects	scores	structure
17	junior	quality	start	years	correlation	structural	level	content	differences	technology
18	academic	lifelong	creative	service	examine	culture	status	textbooks	experimental	labor
19	general	reform	action	conditions	positive	capital	technical	student	concept	based
20	science	policies	interviews	smes	variable	direct	framework	articles	level	sector

<표 계속>

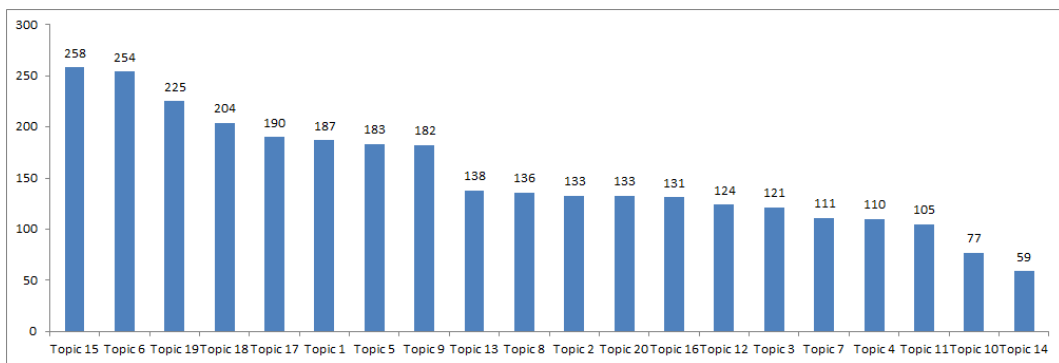
빅데이터 기반의 직업능력개발 연구동향 분석과 시사점 (오호영 외)

	Topic 11 사회정책과 청년고용	Topic 12 직업훈련 프로그램	Topic 13 인적자원개 발 정책	Topic 14 개발 협력	Topic 15 수련과 노동시장	Topic 16 프로그램 평가	Topic 17 역량과 성과관리	Topic 18 온라인· 평생교육	Topic 19 교육공학	Topic 20 진로지도
1	social	training	development	information	workers	evaluation	competency	learning	design	career
2	vet	vocational	human	project	labor	program	hrd	based	instructional	work
3	countries	programs	resources	development	skill	based	performance	strategies	educational	experience
4	policy	development	policy	contents	skills	developed	competencies	learners	technology	support
5	youth	support	government	north	market	items	management	environment	instruction	guidance
6	employment	ncs	cooperation	made	wage	conducted	development	web	ory	program
7	labor	institutes	industry	developing	firms	method	corporate	learner	principles	decision
8	market	system	local	conducted	rate	develop	companies	transfer	computer	level
9	economic	program	regional	basic	firm	criteria	resource	online	approach	making
10	life	courses	policies	south	unemployment	assessment	level	lifelong	media	success
11	aged	institutions	support	year	employment	validity	core	cyber	paper	preparation
12	policies	skills	resource	including	worker	experts	business	feedback	oretical	development
13	lifelong	ability	financial	resources	time	test	human	support	based	life
14	people	participation	universities	expert	small	suggested	key	directed	digital	counseling
15	welfare	unemployed	system	materials	income	finally	company	strategy	context	family
16	capital	order	역량과 성과관리	curriculum	regular	order	expertise	environments	models	ethics
17	participation	standard	organizations	developed	investment	components	corporations	participation	distance	important
18	system	trainees	national	selected	level	indicators	organization	time	virtual	maturity
19	middle	fur	projects	integrated	increase	scale	enterprises	regulated	field	experiences
20	support	improve	central	methods	higher	areas	importance	instructor	ories	care

20개의 주제 중 가장 연구가 많이 된 주제는 ‘숙련과 노동시장(topic 15)’이며, 그 다음은 조직관리(topic 6), 교육공학(topic 19), 온라인·평생교육(topic 18), 역량과 성과관리(topic 17), 학교교육(topic 1), 심리적·정서적 유인(topic 5), 학생(topic 9) 등의 순으로 나타났다. 20개의 주제 중 인력 수요와 공급(topic 10)과 개발협력(topic 14)에 속한 연구물은 각각 77편과 59편으로 가장 적다.

[그림 10] 주제별 연구물 편수

(단위: 편)



주 : Topic1-학교 교육, Topic2-직업교육 제도 및 정책, Topic3-지식·역량, Topic4-일자리·고용, Topic5-심리적·정서적 유인, Topic6-조직관리, Topic7-자격, Topic8-교사, Topic9-학생, Topic10-인력 수요·공급, Topic11-사회정책과 청년고용, Topic12-직업훈련 프로그램, Topic13-인적자원개발 정책, Topic14-개발 협력, Topic15-숙련과 노동시장, Topic16-프로그램 평가, Topic17-역량과 성과관리, Topic18-온라인·평생교육, Topic19-교육공학, Topic20-진로지도

다음은 시간의 흐름에 따라 연구 주제가 어떻게 변화해 왔는지를 살펴보았다.³⁾ 1999년까지는 교육공학(topic 19) 및 온라인·평생교육(topic 18)과 관련된 연구가 가장 활발하게 진행되었으며, 2005년까지 유지되다 점차 쇠퇴하는 추세를 보였다. 『한국노동경제논집』이 분석에 포함된 2000년부터는 숙련과 노동시장(topic 15)에 대한 연구가 급격히 증가하여 현재까지 활발히 진행되고 있다. 2006~2010년에는 『역량과 성과관리 연구』가 창간되면서 역량과 성과관리(topic 17)과 조직관리(topic 6)에 대한 연구가 많이 수행되었으며, 2011~2015년에는 조직관리(topic 6) 분야를 필두로 조직에서의 심리적·정서적 유인(topic 5)에 대한 연구가 활발히 이루어졌다. 2016년 이후에도 여전히 조직관리(topic 6), 조직에

3) 2000년 이후 창간호(예, 직업교육연구, 역량과 성과관리 연구, 한국노동경제논집)가 포함됨으로써 시계열 분석의 해석상 문제가 발생할 수 있으나 이들 학술지를 제외할 경우 직능원 개원 이후 전기간의 분석이 어려운 점을 고려하여 분석에 포함. 시계열 분석결과의 해석에서는 이점을 유의할 필요가 있음.

빅데이터 기반의 직업능력개발 연구동향 분석과 시사점 (오호영 외)

서의 심리적·정서적 유인(topic 5), 숙련과 노동시장(topic 15)과 관련된 연구가 대세를 이루고 있음을 알 수 있다.

직능원의 기본 보고서와 교육훈련 및 고용 분야 학술지가 주로 다루고 있는 연구 주제를 비교해보며 다음과 같은 특징을 발견할 수 있다. 직능원은 인적자원개발정책(topic 13)과 관련된 연구를 가장 많이 수행해왔으며, 그 다음은 자격(topic 7), 직업훈련프로그램(topic 12)에 대한 연구가 주로 진행되어왔다. 그 외에도 직업교육제도 및 정책(topic 2)과 사회정책과 청년고용(topic 11)과 같이 직능원의 고유한 연구 분야에 대한 연구가 타 학술지에 비해 상대적으로 많이 수행되어 왔다.

<표 11> 주제별 연구동향 시계열 비교

(단위: 편)

구분	~1999년	2000~2005년	2006~2010년	2011~2015년	2016년~
Topic 1	10	37	44	58	38
Topic 2	20	25	33	35	20
Topic 3	5	32	24	36	24
Topic 4	5	14	35	32	24
Topic 5	3	8	58	76	38
Topic 6	2	2	59	122	69
Topic 7	11	28	30	34	8
Topic 8	7	15	27	50	37
Topic 9	16	39	56	55	16
Topic 10	9	12	24	21	11
Topic 11	2	21	35	32	15
Topic 12	13	29	30	39	13
Topic 13	3	49	57	17	12
Topic 14	1	18	19	13	8
Topic 15	2	61	59	90	46
Topic 16	6	19	45	44	17
Topic 17	3	22	67	71	27
Topic 18	28	70	38	44	24
Topic 19	85	56	30	34	20
Topic 20	5	17	36	57	18

주 : Topic1-학교 교육, Topic2-직업교육 제도 및 정책, Topic3-지식·역량, Topic4-일자리·고용, Topic5-심리적·정서적 유인, Topic6-조직관리, Topic7-자격, Topic8-교사, Topic9-학생, Topic10-인력 수요·공급, Topic11-사회정책과 청년고용, Topic12-직업훈련 프로그램, Topic13-인적자원개발 정책, Topic14-개발 협력, Topic15-숙련과 노동시장, Topic16-프로그램 평가, Topic17-역량과 성과관리, Topic18-온라인·평생교육, Topic19-교육공학, Topic20-진로지도

학술지에서도 학술지 고유의 세부 전공영역에 따라 주요 연구 주제가 명확하게 구분되었다. 『직업능력개발연구』는 숙련과 노동시장(topic 15), 조직에서의

빅데이터 기반의 직업능력개발 연구동향 분석과 시사점 (오호영 외)

심리적·정서적 유인(topic 5), 학교교육(topic 1)을 주제로 한 논문이 많았으며, 『교육공학연구』는 교육공학(topic 19), 학생(topic 9), 온라인·평생교육(topic 18)에 대한 연구가 주를 이루었다. 『기업과 인재연구』는 조직관리(topic 6), 역량과 성과관리(topic 17), 온라인·평생교육(topic 18)에 대한 연구가 주를 이루었고, 『한국노동경제논집』에서는 숙련과 노동시장(topic 15)에 대한 논문이 집중적으로 실렸다. 『역량과 성과관리 연구』는 조직관리(topic 6)와 역량과 성과관리(topic 17), 프로그램 평가(topic 16)에 대한 연구물이 주로 실렸으나, 『직업교육연구』는 학교교육(topic 1), 진로지도(topic 20), 조직에서의 심리적·정서적 유인(topic 5) 외에도 다양한 연구 주제를 다루고 있다.

<표 12> 학술지 및 직능원 기본보고서의 주요 연구 주제

(단위: 편)

	직능원 기본보고서	직업능력 개발연구	교육공학연구	기업과 인재연구	한국노동 경제논집	역량과 성과관리연 구	직업교육연구
Topic 1	75	20	6	1	7	5	73
Topic 2	76	19	8	3	1	9	17
Topic 3	12	9	52	17	1	13	17
Topic 4	38	14	1	9	15	5	28
Topic 5	-	25	62	14	2	23	57
Topic 6	3	17	7	64	2	108	53
Topic 7	78	8	4	1	2	6	12
Topic 8	10	3	57	6	-	17	43
Topic 9	2	3	149	4	2	3	19
Topic 10	56	5	1	2	5	2	6
Topic 11	67	12	-	1	11	4	10
Topic 12	77	9	2	6	-	9	21
Topic 13	110	9	1	-	-	16	2
Topic 14	42	1	4	-	-	3	9
Topic 15	54	35	-	6	144	4	15
Topic 16	17	11	37	11	-	24	31
Topic 17	28	13	7	47	1	63	31
Topic 18	13	9	137	25	-	9	11
Topic 19	4	4	199	13	-	3	2
Topic 20	41	12	-	3	1	10	66
합계	803	238	734	233	194	336	523

주 : Topic1-학교 교육, Topic2-직업교육 제도 및 정책, Topic3-지식·역량, Topic4-일자리·고용, Topic5-심리적·정서적 유인, Topic6-조직관리, Topic7-자격, Topic8-교사, Topic9-학생, Topic10-인력 수요·공급, Topic11-사회정책과 청년고용, Topic12-직업훈련 프로그램, Topic13-인적자원개발 정책, Topic14-개발 협력, Topic15-숙련과 노동시장, Topic16-프로그램 평가, Topic17-역량과 성과관리, Topic18-온라인·평생교육, Topic19-교육공학, Topic20-진로지도

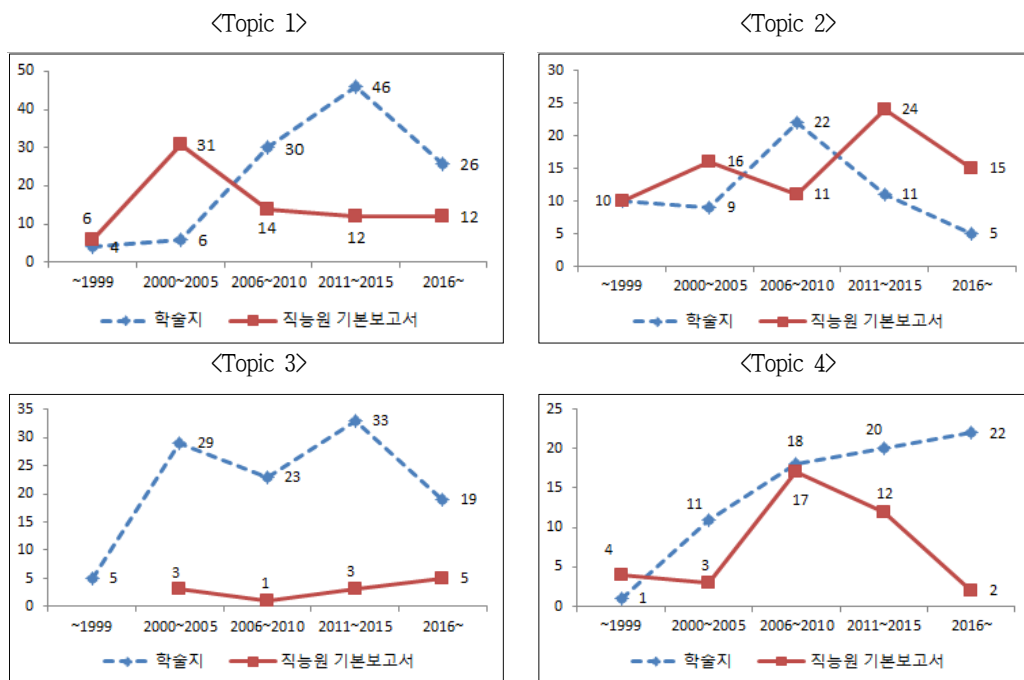
직능원과 교육훈련·고용 분야 학술지의 연구 동향을 시계열로 살펴보면 다음과 같다. 학교교육(topic 1)과 관련한 주제의 경우, 직능원은 2000대 초반에 연구가 활발히 이루어졌다가 2000년대 중후반부터 급격히 감소한 반면, 교육훈련·고용 분야의 학술지는 2000년대 중후반부터 논문 수가 급격히 증가하여 2010년 초반에 정점을 찍고 2016년 이후부터 줄어드는 추세를 보였다. 직업교육제도 및 정책(topic 2)은 시기별로 변동은 있으나 대체로 증가하는 추세인 반면, 학술지의 경우 2000년대 중후반에 정점을 찍고 급감하는 추세를 보이고 있다. 지식·역량(topic 3)은 직능원에서 거의 다루지 않는 연구 주제이나, 학술지에서는 2000년 이후부터 활발히 연구되고 있는 주제이다. 일자리·고용(topic 4)에 대한 주제는 직능원의 경우 2006~2010년에 정점을 찍은 후 급격히 줄어들었으나 학술지에서는 꾸준히 증가하고 있는 추세이다. 심리적·정서적 유인(topic 5)과 조직관리(topic 6)의 경우 직능원의 기본연구에서 거의 다루지 않는 주제이나 학술지에서는 2006년부터 관련 논문이 급격히 증가하고 있다. 자격(topic 7)은 직능원의 특징적인 연구 주제로, 특히 2011~2015년에 연구가 활발히 이루어졌다. 학술지에서는 2006~2010년까지는 꾸준히 증가하고 있으나 2011년 이후로는 빠르게 줄어들고 있는 추세이다. 교사(topic 8)과 학생(topic 9)은 직능원에서 거의 수행되고 있지 않는 연구 주제인 반면, 학술지에서는 2015년까지 관련 논문이 지속적으로 증가해왔으며 논문 수도 다른 주제에 비해 많은 편에 속한다. 인력 수요·공급(topic 10)은 직능원과 학술지 간의 연구 추이는 비슷하나 직능원에서 좀 더 많은 연구가 이루어져 왔음을 알 수 있다.

사회정책과 청년고용(topic 11)의 경우 학술지는 지속적으로 관련 논문이 줄어들고 있는 반면, 직능원은 개원한 이후부터 2015년까지 관련 분야의 연구가 꾸준히 증가하다 최근에 약간 줄어드는 추세를 보이고 있다. 직업훈련 프로그램(topic 12) 분야에서는 2000년 초반과 2010년 초반에 직능원에서 활발한 연구가 이루어졌으나 2016년 이후에는 크게 감소하는 추세를 보인다. 학술지의 경우 직업훈련 프로그램(topic 12)을 주제로 한 연구물이 2015년까지 꾸준히 증가해오다 2016년 이후부터는 다소 줄어드는 추세이다. 인적자원개발정책(topic 13) 연구주제의 경우 직능원에서는 2000~2010년까지 44편과 39편의 보고서를 발간하며 활발히 연구되다 2011년 이후부터는 감소하였고, 학술지에서도 마찬가지로 2011년 이후로 연구물이 감소하는 경향을 보였다. 개발협력(topic 14) 주제에 대해서는 직능원의 경우 2000~2005년 동안 16편의 기본보고서를 발간한 이후

관련 연구가 지속적으로 감소해왔다. 학술지에서도 개발협력(topic 14)과 관련된 연구가 많지 않아 교육훈련·고용 분야에서는 관심을 받는 연구 주제는 아닌 것으로 판단된다. 숙련과 노동시장(topic 15)은 직능원에서 꾸준히 연구를 진행해 온 주제이다. 학술지에서도 2000년 이후 많은 연구가 진행되어 왔으며, 특히 2011~2015년에는 교육훈련·고용 분야 학술지에서 72편의 논문이 발표되어 정점을 찍었다. 2016년 이후에는 다소 줄어들었으나 여전히 다른 연구 주제에 비해 많은 관심을 받고 있다. 프로그램 평가(topic 16)와 역량과 성과관리(topic 17)의 경우 학술지에서 2015년까지 활발히 연구되던 주제인 반면, 직능원에서는 관련 연구가 많이 수행되지 않았다. 온라인·평생교육(topic 18)을 주제로 한 논문은 2000년 초반을 정점으로 감소하는 추세이며 직능원에서도 근근이 연구의 맥을 이어오다 최근에는 관련 연구가 거의 이루어지지 않고 있다. 교육공학(topic 19) 또한 2000년대부터 관련 연구가 급감하는 추세이며 직능원에서는 거의 다루지 않은 연구 주제이다. 진로지도(topic 20)은 그 수는 많지 않으나 직능원에서 꾸준히 연구되어오고 있는 주제로 2011~2015년에 가장 많은 보고서가 발간되었다. 학술지에서는 2000년 이후 급격히 관련 논문 수가 증가하다 2011~2015년에 정점을 찍고 감소하는 추세를 보인다.

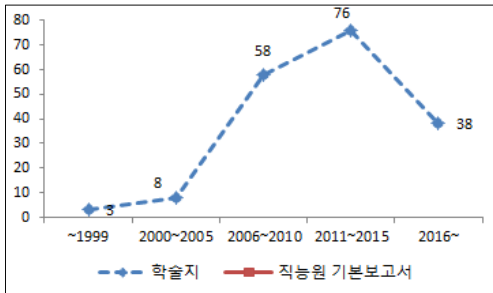
[그림 11] 직능원 기본보고서와 학술지 연구 동향 시계열 비교

(단위: 편)

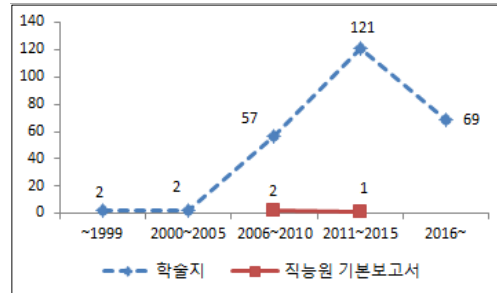


빅데이터 기반의 직업능력개발 연구동향 분석과 시사점 (오호영 외)

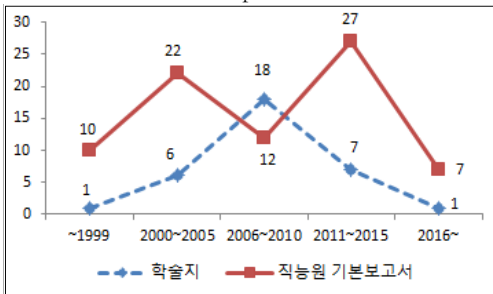
<Topic 5>



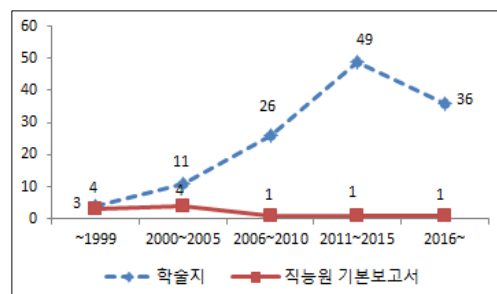
<Topic 6>



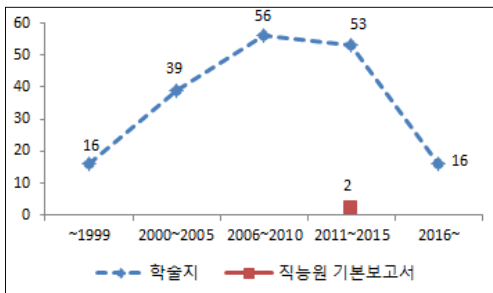
<Topic 7>



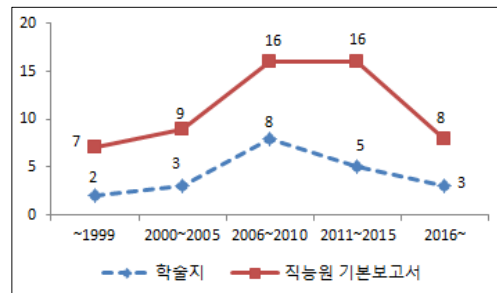
<Topic 8>



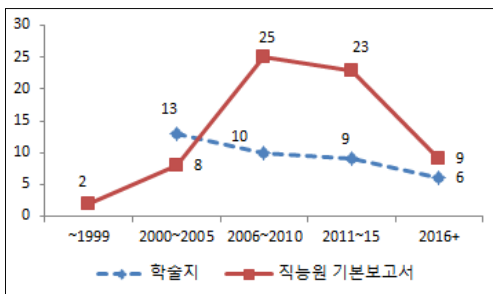
<Topic 9>



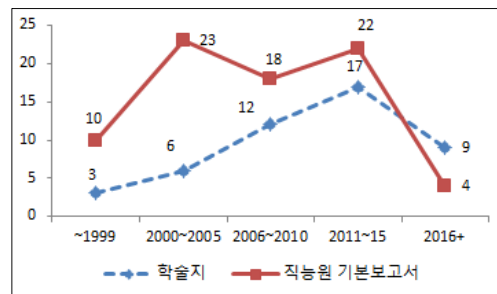
<Topic 10>



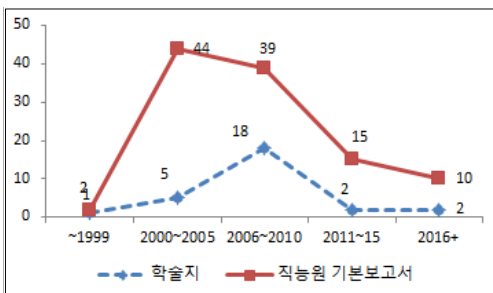
<Topic 11>



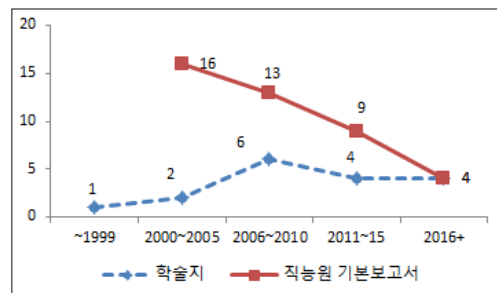
<Topic 12>

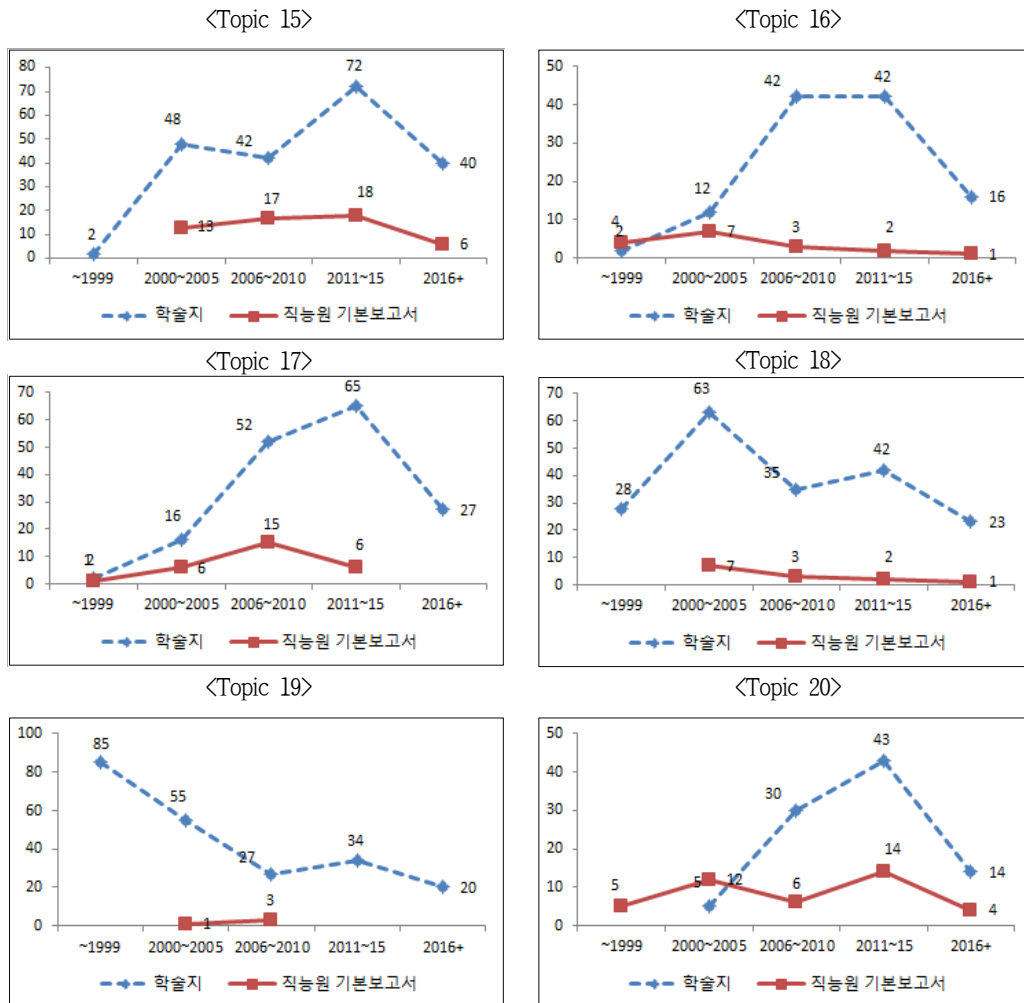


<Topic 13>



<Topic 14>





주 : Topic1-학교 교육, Topic2-직업교육 제도 및 정책, Topic3-지식·역량, Topic4-일자리·고용, Topic5-심리적·정서적 유인, Topic6-조직관리, Topic7-자격, Topic8-교사, Topic9-학생, Topic10-인력 수요·공급, Topic11-사회정책과 청년고용, Topic12-직업훈련 프로그램, Topic13-인적자원개발 정책, Topic14-개발 협력, Topic15-숙련과 노동시장, Topic16-프로그램 평가, Topic17-역량과 성과관리, Topic18-온라인·평생교육, Topic19-교육공학, Topic20-진로지도,

IV. 결론 및 시사점

직능원과 교육훈련·고용 분야 학술지의 1990년 이후 연구 동향을 주제별, 시계열로 살펴보면 다음과 같은 특징을 찾을 수 있다.

첫째, 직능원은 인적자원개발정책(topic 13)과 관련된 연구를 가장 많이 수행

해왔으며, 그 다음은 자격(topic 7), 직업훈련프로그램(topic 12)에 대한 연구가 이루어졌다. 그 외에도 직업교육제도 및 정책(topic 2)과 사회정책과 청년고용(topic 11)과 같이 직능원의 고유한 연구 분야에 대한 연구가 타 학술지에 비해 상대적으로 많이 수행되어 왔다. 학술지의 연구동향을 보면 20개의 주제 중 가장 연구가 많이 된 주제는 ‘숙련과 노동시장(topic 15)’이며, 그 다음은 조직관리(topic 6), 교육공학(topic 19), 온라인·평생교육(topic 18), 역량과 성과관리(topic 17), 학교교육(topic 1), 심리적·정서적 유인(topic 5), 학생(topic 9) 등의 순으로 나타났다.

둘째, 시계열적으로 살펴보면 학교교육(topic 1)과 관련한 주제의 경우, 직능원은 2000대 초반에 연구가 활발히 이루어졌다가 2000년대 중후반부터 급격히 감소한 반면, 교육훈련·고용 분야의 학술지는 2000년대 중후반부터 논문 수가 급격히 증가하여 2010년 초반에 정점을 찍고 2016년 이후부터 줄어드는 추세를 보였다. 직업교육제도 및 정책(topic 2)은 시기별로 변동은 있으나 대체로 증가하는 추세인 반면, 학술지의 경우 2000년대 중후반에 정점을 찍고 급감하는 추세를 보이고 있다. 자격(topic 7)은 직능원의 특징적인 연구 주제로, 특히 2011~2015년에 연구가 활발히 이루어졌다. 학술지에서는 2006~2010년까지는 꾸준히 증가하고 있으나 2011년 이후로는 빠르게 줄어들고 있는 추세이다. 직업훈련 프로그램(topic 12) 분야에서는 2000년 초반과 2010년 초반에 직능원에서 연구를 활발히 수행하였으나 2016년 이후에는 크게 감소하는 추세를 보인다. 학술지의 경우 직업교육훈련 프로그램(topic 12)을 주제로 한 연구물이 2015년까지 꾸준히 증가해오다 2016년 이후부터는 다소 줄어드는 추세이다. 인적자원개발정책(topic 13) 연구주제의 경우 직능원에서는 2000~2010년까지 44편과 39편의 보고서를 발간하며 활발히 연구되다 2011년 이후부터는 감소하였고, 학술지에서도 마찬가지로 2011년 이후로 연구물이 감소하는 경향을 보였다.

이상의 분석을 통하여 시사점을 도출해보면 다음과 같다.

첫째, 직능원 기본연구의 연구동향과 유관학술지들의 연구동향은 뚜렷한 차이점을 보이고 있으며 이는 분석대상 학술지의 특성, 국책연구기관으로서 직능원의 사회적 책무성과 연구지향에서의 차이를 반영하는 것으로 해석된다. 예를 들어 직능원에서는 인적자원개발정책(Topic 13)에서 2000~2010년의 기간 중 압도적인 비중의 연구를 수행하였는데, 이는 국내연구가 상대적으로 부족한 상황에서 직능원이 관련 학술연구와 정책연구를 선도적으로 수행해온 결과에 따른

것으로 평가된다. 직능원이 당시 정부부처를 지원하기 위하여 ‘인적자원개발 지원센터’로 지정되었던 시대적 상황과 무관하지 않으며 정책연구를 통해 정부정책을 선도 및 지원하고 사회적 관심을 모으는 구심점 역할을 적극적으로 수행한 결과로 해석된다.

둘째, 직능원의 정관상 설립목적, 정부의 정책과 직결된다고 할 수 있는 주제에 대한 연구가 충실이 이루어져 국책연구기관으로서 책무성을 충실히 이행한 것으로 평가된다. 직업교육 제도 및 정책(Topic 2), 자격(Topic 7), 사회정책과 청년고용(Topic 11), 직업훈련 프로그램(Topic 12) 등은 개원 이래 지속적으로 연구가 수행되어 왔다.

셋째, 학술지에서 활발히 연구가 수행되고 있음에도 불구하고 직능원에서 감소하거나, 부진한 연구주제를 면밀히 검토하여 연구필요성을 고민해 볼 필요가 있다. ‘Topic4-일자리·고용’은 국가적 현안이 되고 있는 일자리 문제가 갖는 중요성과 직능원의 설립목적이나 책무성에 비추어 관심을 가지고 지속적으로 연구해야 할 주제임에도 최근 연구가 크게 위축되어 시급히 개선할 필요가 있다. ‘Topic8-교사’는 교육에서 교사가 차지하는 중요성에 비추어 연구가 보다 활성화될 필요성이 있음에도 불구하고 개원초기에 비교하여 최근에는 연구성과가 미미하여 좀 더 많은 연구관심이 필요해 보인다. ‘Topic9-학생’ 역시 연구 주제로서의 중요성이나 정책적인 요구에 비추어 연구가 거의 이루어지지 않아 좀 더 많은 관심이 요청되며, Topic18-온라인·평생교육도 최근의 온라인학습 확산, 평생학습의 중요성 추세에 비추어 연구필요성이 높아 보인다. 직능원 연구역량의 분야별, 이슈별, 주제별 적정한 배분과 더불어 외부의 연구 트렌드를 고려한 체계적이고 조직적인 연구주제의 발굴 및 수행노력을 통해 연구주제의 다양성을 제고하고 여러 분야에서 제기되고 있는 다양한 정책연구니즈를 어떻게 충족할 것인지 고민이 필요하다.

이상의 분석결과를 전체적으로 종합해보면 직능원의 연구는 학계의 연구동향과 일정정도의 차별성을 갖고 있으며 이는 본 연구에서 고려한 분석대상 학술지가 제한적이라는 한계와 더불어 직능원이 국책연구기관으로서 요청받고 있는 책무성에 상당부분 기인하는 것으로 볼 수 있다. 즉, 직능원 정관상의 기능과 역할, 그리고 이에 따른 연구부서 조직구성, 연구자의 전공구성 등이 좀 더 다양한 연구주제 선정을 제약하는 요인으로 작용하였을 것이다. 향후 직능원의 연구주제 발굴 및 선정에 있어서 학계 연구동향에 대한 면밀한 분석과 평가결

과를 바탕으로 연구주제의 다양성을 확보하고 연구원 외부의 연구동향을 좀 더 적극적으로 반영하는 체계적 노력이 필요해 보인다.

본 연구는 다음과 같은 한계점을 갖는다. 우선, 좀 더 정치한 단어사전을 작성하고 이를 토대로 불필요한 단어들을 최대한 통제된 상태에서 텍스트마이닝을 시도했어야 하나, 연구기간의 제약으로 인하여 충실한 단어사전의 작성이 미흡한 점을 지적할 수 있다. 이는 본 연구의 연장선에서 추가적인 연구를 통해 직업능력개발 분야의 단어사전을 작성하는 축적의 시간을 통해 장기적으로 극복해야 할 과제이다. 둘째, 관련되는 이슈로서 직업능력개발 관련 유의미한 단어의 추출과 이를 기반으로 한 토픽모델링을 통해 토픽간의 차별성과 정확한 토픽의 식별 노력이 강화될 필요성이다. 현재 논문에서는 20개의 토픽을 기반으로 연구주제의 특징과 변화 동향을 파악하고 있는데, 이러한 토픽분석이 적절한 것인지, 연구진이 작성한 토픽의 명칭이 타당한 것인지, 토픽간의 관련성은 무엇인지, 중심토픽은 무엇인지 등에 대한 추가적인 분석을 통해 연구방향성 정립에 좀 더 유의미한 정보추출이 가능할 것으로 보인다. 마지막으로 본 연구에서는 6개의 유관 학술지만을 대상으로 분석을 시도하였으나, 범위를 좀 더 넓혀 국제학술지의 연구동향과의 비교도 우리에게 시사하는 바가 클 것으로 생각되므로 연구외연을 확장할 필요성도 있다.

참고문헌

- 김상겸(2016). 『토픽모델링을 이용한 국내 산업공학 연구동향 분석』. 석사학위 논문. 서울과학기술대학교.
- 김성근, 조혁준, 강주영(2016). 학술연구에서의 텍스트 마이닝 활용 현황 및 주요 분석기법. *정보화연구(구 정보기술아키텍처연구)*, 13(2), 317-329.
- 김영우(2017). 『Do it! 쉽게 배우는 R 데이터 분석』. 이지스퍼블리싱.
- 김은지, 황영아, 송영수(2017). 직업교육연구 학회지의 연구동향 분석: 내용분석 및 네트워크 분석의 활용. *직업교육연구*, 38(6), 1-27.
- 김학준(2015). 언어네트워크 분석을 활용한 국내관광 분야 연구 동향 분석. *관광레저연구*, 27(6), 191-210.
- 백영민(2017). 『R을 이용한 텍스트 마이닝』. 한올아카데미.
- 성보경, 유연우(2018). 빅데이터 기법을 활용한 직업훈련 요구분석. *한국융합학회 논문지*, 9(5), 21-26.
- 성지은, 박기량(2014). 빅데이터를 활용한 정책 사례 분석과 시사점. *과학기술정책*, 24(2), 94-106.
- 신배정, 장희순(2014). ‘부동산’ 용어의 의미연결망 분석. *주거환경*, 12(1), 101-113.
- 유예림(2017). 빅데이터 분석 기법을 활용한 2015 개정 교육과정 정책에 대한 언론보도 분석. *박사학위논문*, 서울대학교.
- 유진은(2016). 데이터 마이닝 기법을 통한 교육 패널데이터 분석: 별점회귀모형과 kyp 자료.
- 이시연, 문혜정(2017). 세계 교통관련 연구의 빅데이터 분석을 통한 교통연구 동향 분석. *한국경영학회 통합학술발표논문집*, , 299-313.
- 장세은, 이수호(2016). 키워드 네트워크 분석을 통한 세계 해운경제의 연구 주제와 동향에 대한 연구. *한국항만경제학회지*, 32(1), 79-95.
- 정의섭, 여운동, 조대연, 서일원(2008). 과학계량학적 정보분석을 통한 건설교통 분야의 유망연구영역도출. *한국콘텐츠학회논문지*, 8(2), 231-238.

진희정, 이재철, 장은수(2014). 네트워크 분석을 통한 국내 변증 연구의 동향.
한국콘텐츠학회논문지, 14(12), 1037-1046.

임동훈(2016). 『R을 이용한 빅데이터 분석』. 자유아카데미.

Andrea Cirillo(2017). R Data Mining. Packt.

Blei, D. M.(2012). Probabilistic Topic Models, *Communications of the ACM*, 55(4)

Julia Silge & David Robinson(2017). Text Mining with R. O' REILLY.