

토픽모델링과 시계열회귀분석을 활용한 정보시스템분야 연구동향 분석

김창식¹ · 최수정¹ · 광기영^{2*}¹국민대학교 비즈니스IT전문대학원²국민대학교 경영대학/비즈니스IT전문대학원

Investigation of Research Trends in Information Systems Domain Using Topic Modeling and Time Series Regression Analysis

Chang-Sik Kim¹ · Su-Jung Choi¹ · Kee-Young Kwahk^{2*}¹Graduate School of Business IT, Kookmin University, Seoul 02707, Korea²College of Business Administration/Graduate School of Business IT, Kookmin University, Seoul 02707, Korea

[요약]

본 연구의 목적은 국내에서 2002년부터 2016년까지 출판된, 대표적인 정보시스템분야 저널의 연구동향을 조사하는 것이다. 연구의 목적을 달성하기 위해서 Asia Pacific Journal of Information Systems, Information Systems Review, The Journal of Information Systems에 출판된 논문의 초록 1,245편을 분석 하였다. 본 연구에서는 최근 중요하게 다루어지는 토픽모델링과 시계열회귀분석 기법을 활용하였다. 토픽모델링 분석결과, 20개의 토픽이 도출되었고 “시스템구축”, “혁신역량”, 및 “고객충성도” 등의 순으로 확인되었다. 둘째, 시계열회귀분석 결과, 상승 추세를 나타내는 토픽으로는 “고객충성도”, “소통혁신”, “정보보호”, 및 “개인정보보호” 가 나타났고, 하락 추세를 나타내는 토픽으로는 “시스템구축” 및 “웹사이트” 가 도출되었다.

[Abstract]

The objective of this study is to examine the trends in information systems research. The abstracts of 1,245 articles were extracted from three leading Korean journals published between 2002 and 2016: Asia Pacific Journal of Information Systems, Information Systems Review, and The Journal of Information Systems. Time series analysis and topic modeling methods were implemented. The topic modeling results showed that the research topics were mainly "systems implementation", "communication innovation", and "customer loyalty". The time series regression results indicated that "customer satisfaction", "communication innovation", "information security", and "personal privacy" were hot topics, and on the other hand, "system implementation" and "web site" were the least popular. This study also provided suggestions for future research.

색인어 : 정보시스템연구동향, 토픽모델링, 시계열회귀분석

Key word : Information System Research Trend, Topic Modeling, Time Series Regression Analysis

<http://dx.doi.org/10.9728/dcs.2017.18.6.1143>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 01 September 2017; Revised 13 October 2017

Accepted 25 October 2017

*Corresponding Author; Kee-Young Kwahk

Tel: +82-02-910-4738

E-mail: kykwahk@kookmin.ac.kr

I. 서론

최근 텍스트분석기법인 토픽모델링과 언어네트워크분석을 이용한 연구들이 다양하게 이루어지고 있다. 문헌정보학 분야의 주요저널을 대상으로 연구주제와 방법을 토픽모델링으로 분석한 연구가 수행되었으며[1], 컨벤션(MICE)분야에서도 연구동향 분석에 언어네트워크분석을 이용하였다[2]. 또한 텍스트마이닝과 소셜네트워크 분석을 적용하여 호텔분야의 연구동향을 분석한 연구도 이루어졌다[3]. 이러한 연구들은 텍스트 분석을 통해 의미를 찾고자 하는 대표적인 문제해결방법이다.

텍스트마이닝을 이용해 연구 주제와 연구 방법을 확인하기 위해서는 동시 출현 단어에 가중치를 부여한다. 이를 통해 연구 주제와 연구 방법의 군집도 파악할 수 있다. 그러나 연구 주제와 연구 방법을 식별하고 이름을 부여하는 것은 전문가의 전문성에 따라 달라질 수 있다는 한계점이 있다[1]. 국내의 다양한 학문 영역에서 텍스트마이닝과 언어네트워크분석기법이 적용되어 연구동향에 대한 트렌드 분석이 이루어지고 있다. 연도별 트렌드의 증가와 하락을 파악하는 일부 연구에서 추세를 토대로 저자가 임의로 해석을 하는 경우가 종종 있다. 이는 텍스트마이닝과 소셜네트워크분석 기법을 활용한 연구가 초기단계라 시계열회귀분석을 감안하지 않았기 때문이라 판단된다. 이러한 문제는 토픽모델링과 시계열회귀분석 기법의 통합적 접근으로 해결 가능하다. 위와 같이 다양한 분야에서 텍스트마이닝 등을 적용한 연구가 수행되었으나, 국내 정보시스템 분야에서 토픽모델링과 시계열분석 기법의 통합관점으로 접근한 연구동향 분석은 미흡하며, 연구가 필요한 상태이다.

본 연구의 목적은 국내 정보시스템 분야의 연구동향을 분석을 위해 최근 주목받는 토픽모델링과 시계열분석기법을 이용하여 파악하는 것이다. 국내 정보시스템 분야 주요저널 3곳의 15년간(2002년~2016년) 게재된 연구의 초록을 분석대상으로 설정하였다. 본 연구의 결과는 다양한 분야의 연구자들이 동향 연구를 수행함에 있어, 필요한 방법론적인 접근법과, 국내 정보시스템 연구자들에게 현재 중요한 분야 또는 중요성이 간과되어 있는 분야에 도움을 줄 수 있다. 이는 정보시스템 연구자들이 향후 연구방향 수립을 하는데 시사점을 제시해 줄 수 있을 것이다.

II. 문헌 연구

2-1 토픽모델링 선행연구

토픽 모델링(Topic Modeling)은 텍스트 내에서 문서의 주제를 도출하기 위해 사용하는 방법론이다[4]-[6]. 이러한 토픽 모델링은 텍스트분석에 자주 활용되고 있다[3], [7]. 토픽 분석에 사용되는 텍스트는 소셜 미디어, 뉴스 데이터, 블로그, 논문 등의 다양한 원천에서 취득할 수 있다[3]. 토픽모델링에서 주로 사용되는 알고리즘은 잠재의미분석(Latent semantic analysis:

LSA)과 잠재디리클레할당(Latent Dirichlet Allocation: LDA)으로 구분할 수 있다[8].

정보검색 경우에 잠재의미색인(Latent semantic indexing: LSI)이라고도 불리는 잠재의미분석은 텍스트 자료의 의미구조를 추출하고, 나타내는데 적합한 수학적이고 통계적인 방법으로, 질의에 의해서 제공되는 용어를 기반으로 문서를 자동으로 색인하고 검색하기 위해서, 차원축소 도구로서 특이값 분해(Singular Value Decomposition: SVD)를 적용한다[9]-[12]. 즉 잠재의미색인 구현의 세부적인 절차에서, 특이값 분해라 불리는 고유벡터 분해 및 요인 분석과 관련된 선형대수 기술이 적용된다[11]-[13].

잠재의미분석의 세부적인 절차는 3단계로 구성되어 진다. ① 잠재의미분석 전처리 단계(Pre-LSA): 텍스트를 계량화하는 단계로서, 원시 데이터의 전처리 작업을 수행한다. ② 잠재의미분석 단계(LSA): 용어를 축소하여 토픽을 도출하는 단계로서, 선택된 용어의 직접 및 간접적 의미 유사성을 계산하고, 유사도 값을 기반으로 그룹화 작업을 한다. ③ 잠재의미분석 사후 단계(Post LSA): 정량분석 단계로서, 토픽-문서 매트릭스를 사용하여 의미 있는 토픽을 추출하고, 이후 분석에서 활용한다[14].

잠재디리클레할당(Latent Dirichlet Allocation, LDA) 알고리즘은 Blei et al.(2003)[15]이 제안한 것으로, 주제 별 단어 수 분포를 대상으로, 주어진 문서의 단어 수를 분석하고, 어떤 주제를 함께 다루는지를 예측한다[6]. 잠재디리클레할당은 R 등의 오픈 소프트웨어를 활용한 토픽모델링에 적용되고 있으며, 최근 많은 연구자들이 사용하고 있다.

본 연구는 잠재의미분석 및 잠재디리클레할당 알고리즘 중 특이값 분해 개념을 바탕으로 하는 잠재의미분석기법이 적용된 SAS Enterprise Miner 14.1(이하 SAS EM)를 활용하였고, <그림 1>은 잠재의미분석의 절차이다[14].

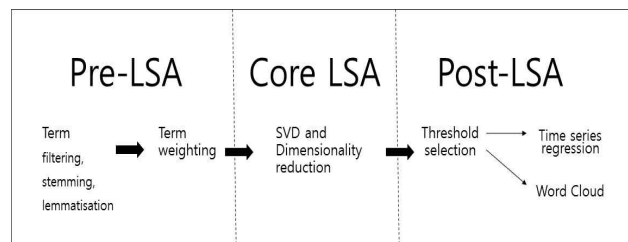


그림 1. 잠재의미분석 절차
Fig. 1. Process of latent semantic analysis

2-2 연구동향 선행연구

국내에서 학술논문을 제공하는 중요한 사이트 중의 한곳에서 ‘연구동향’이라는 키워드로 검색한 결과 1969년부터 2017년까지 총 4,468건의 논문이 검색되었다. 세부적으로, 공학 1,745건, 사회과학 1,123건, 인문학 483건, 자연과학 462건, 예술체육

199건, 농수해양학 160건, 의약학 119건, 복합학 70건 등으로 구성되어 있다. 또한 사회과학은 경제/경영학 212건, 교육학 168건, 사회과학일반 125건, 문헌정보학 119건, 사회/사회복지학 95건, 행정학 82건 등으로 나타났다(www.dbpia.com).

최근의 경영/경제학분야 연구동향과 관련 주요논문으로는 한국경영학회에서 발간하는 학술지 Korea Business Review(KBR)의 454편을 대상으로 계량서지학적 관점 분석한 연구[16], 국내외 ‘전략적 제휴’ 키워드로 검색된 국내 논문 170편과 국외 논문 179편을 대상으로 제휴의 동기, 유형, 이론, 성과에 초점을 둔 연구[17], 국내 한국교육학술정보원 홈페이지에서 ‘교육서비스 품질, 고객만족, 학생만족’이라는 키워드로 검색된 학술연구논문과 박사학위논문을 대상으로 한 연구[18], 소셜네트워크분석을 활용한 한국무역학회지 공저자 네트워크 분석 연구[19], 계량서지학적 분석을 적용하여 서비스경영학회 608편을 대상으로 저자유형, 연구주제, 연구방법, 연구대상 산업분야 등을 분석한 연구[20] 등이 있다.

텍스트분석 기법 중 언어네트워크분석을 기반으로 수행된 연구로는, MICE[2], 언론학[21], 민간경비[22] 등이 있다. 또한 텍스트마이닝 기법을 기반으로 암반공학분야 SCI논문의 주제어 분석[23], 문헌정보학분야 논문 145건을 대상으로 문헌 클러스터링과 유사도 네트워크 분석[24], 텍스트마이닝과 소셜네트워크분석 기법을 활용한 호텔분야 연구동향 분석[3] 등도 수행되었다. 토픽모델링과 시계열회귀분석 기법을 적용한 핀테크 기술동향[25] 연구에서는 각 토픽의 유망기술(Hot Topic)과 쇠퇴기술(Cold Topic)을 통계적 검정을 통해 도출하기도 하였다.

III. 연구 방법

3-1 분석 대상

정보시스템분야 국내 연구의 동향을 분석하기 위해서, 웹오브사이언스(Web of Science) 데이터베이스에서, 3개의 정보시스템 국내저널(Asia Pacific Journal of Information Systems: APJIS, Information Systems Review: ISR, The Journal of Information Systems: JIS)을 대상으로 2002년부터 2016년까지 1,245편 논문의 초록을 분석대상으로 선정하였다.

국내 정보시스템 선도저널 3개의 연도별/저널별 논문 수 분석결과, 2002년도 78편에서 상승하는 추세를 보이고 있으나, 2006년도 105편, 2008년도 103편을 정점으로 하락하는 형태로 나타났다. 2013년은 58편으로 최저치를 보이고 있으며, 2015년 84편, 2016년도 76편으로 2000년대 초반 수준으로 회복되었다. <표 1>은 분석대상 논문 편수이고, <그림 2>는 분석대상 논문의 연도별 발행 추세를 나타낸다.

표 1. 분석 대상 논문 편수

Table 1. Article Number of Analysis

	2002-2006	2007-2011	2012-2016	Total
APJIS	191	129	115	435
ISR	142	156	100	398
JIS	95	173	144	412
Total	428	458	359	1,245

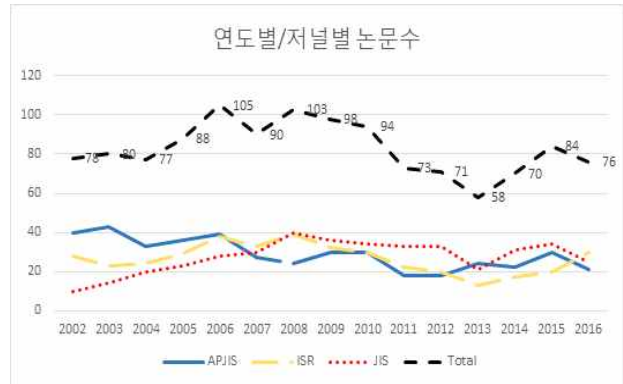


그림 2. 연도별/저널별 논문수

Fig. 2. Number of articles per year/journal

3-2 분석방법 및 절차

본 연구는 1,245편의 논문 초록을 대상으로 토픽모델링과 시계열회귀분석을 통해 기간별 연구동향을 파악하였다. 분석은 전처리, 토픽모델링, 시계열회귀분석 단계로 수행하였다. 첫째, 연구대상 논문의 분석을 위한 전처리 작업에는 Excel을 주로 활용하였다. Excel에서 전처리 완료된 데이터는 SAS Enterprise Guide 7.2를 통하여 SAS 분석용 파일로 변환하였다. 둘째, SAS EM을 활용한 토픽모델링은 ① 분석용 텍스트데이터 불러오기 ② 텍스트 파싱 ③ 텍스트 필터 ④ 토픽모델링의 절차의 순으로 진행된다[26]. 셋째, 토픽모델링의 결과 임계치 값(Cutoff value)을 연도별로 합당한 값 기준으로 SPSS를 활용하여, 시계열회귀분석을 진행하였다.

<그림 3>은 SAS EM의 토픽모델링 프로세스이고, <표 2>는 SAS EM의 토픽모델링 결과 샘플이다.

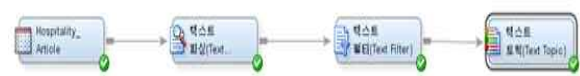


그림 3. 토픽모델링 프로세스

Fig. 3. Process of Topic Modeling

표 2. 토픽 모델링 결과 샘플

Table 2. Result of Topic Modeling(Topic-Document Matrix)

Year	Document number	Weight				Cutoff value			
		t1	t2	...	t20	T1	T2	...	T20
2002	D1	0.040	-0.058	...	-0.001	0	0	...	0
2002	D2	0.006	0.127	...	-0.042	0	1	...	0
2002	D3	0.058	0.162	...	-0.004	0	1	...	0
...
2016	D1245	0.014	0.018	...	0.035	0	0	...	0

IV. 분석결과

4-1 토픽모델링 분석

본 연구에서의 토픽 모델링은 SAS EM을 활용하였고, 워드클라우드에는 R을 활용하였다. 토픽 모델링의 빈도는 각 토픽이 얼마나 많이 등장하였는지를 나타내주며, 토픽은 핵심 키워드를 대표하는 용어로 정의하였다.

워드클라우드는 텍스트를 시각적으로 나타낼 수 있는 대표적인 방법 중의 하나이다. 워드클라우드를 통해 토픽모델링의 결과를 좀 더 직관적으로 표현할 수 있다. 워드클라우드는 각 토픽의 문서 빈도를 기준으로 R을 통해 구현하였다. 정보시스템과 관련하여 시스템구축 토픽이 가장 많이 나타난 반면에, 개인정보 토픽이 가장 적게 나타났다. <그림 4>는 상기 토픽 모델링 결과를 워드클라우드를 통해 나타낸 것이다.

<표 3>은 2002년에서 2016년까지의 토픽 모델링 결과를 나타낸 것이다. 이 기간 논문에서는 시스템구축, 혁신역량, 서비스전략, 고객충성도, 서비스전략 등의 순으로 주제가 나타났다. 시스템구축이 가장 높은 순위에 위치한 것을 볼 때, 정보시스템 분야의 중요한 연구 주제는 성공적인 시스템 구축이라고 판단할 수 있다[27][28].



그림 4. 워드클라우드 결과
Fig. 4. Result of Word Cloud

표 3. 토픽모델링 결과

Table 3. Result of Topic Modeling (2002~2016)

Topic	Keyword	Freq	
T10	시스템구축	system, business, implementation, success, information system	185
T02	혁신역량	innovation, performance, capability, firm, adoption	170
T17	고객충성도	customer, loyalty, customer satisfaction, satisfaction, service	167
T18	서비스전략	market, industry, game, strategy, service	167
T20	소통혁신	journal, media, network, communication, innovation	163
T19	고객만족	user, continuance, intention, behavior, satisfaction	149
T08	서비스품질	quality, satisfaction, service, user satisfaction, user	139
T07	e비즈니스신뢰	trust, mall, internet, internet shop, shop mall	136
T03	제품리뷰	product, consumer, purchase, price, review	132
T14	프로젝트교육	learning, capability, learner, project, education	123
T01	기술수용모델	acceptance, usefulness, ease, tam, intention	121
T11	지식경영	knowledge, knowledge management, management, culture, organization	118
T09	프로젝트리스크	project, risk, software, development, management	110
T04	예측기술	rating, prediction, credit, algorithm, technique	100
T06	가상커뮤니티	community, virtual community, loyalty, online community, group	95
T15	아웃소싱역량	outsourcing, capability, vendor, client, service	93
T13	웹사이트	web, site, web site, user, usability	88
T16	공급망관리	supply, chain, supply chain, supplier, integration	76
T05	정보보안	security, information security, information, awareness, risk	64
T12	개인정보	privacy, concern, personal information, service, user	53

4-2 시계열회귀분석

시간에 따른 정보시스템 분야의 핵심 토픽의 변화를 파악하기 위하여, 토픽모델링 결과 임계치 값(Cutoff value)을 연도별로 합산한 값을 기준으로, SPSS를 활용하여 시계열회귀분석을 하였다. 시계열회귀분석 결과를 통해 각 토픽들의 15년간 연도별 추세를 파악할 수 있다. 또한 시계열회귀분석의 유의확률 값이 통계적으로 유의미하면서, 회귀계수 값을 기준으로 양수(+)로 나타나는 토픽을 핫 토픽(Hot Topic), 음수(-)로 나타나는 토픽을 콜드 토픽(Cold Topic)으로 구분할 수 있다[29]. 핫 토픽은 상승추세를 의미하고, 콜드 토픽은 하향 추세를 의미한다. 유의하지 않은 토픽은 중립토픽(Neutral Topic)을 의미한다.

<표 4>은 시계열회귀분석 결과로서, 통계적으로 유의미한 상승 토픽(Hot), 하락 토픽(Cold), 중립토픽(-)을 나타낸다. 또한 <그림 5>는 연도별 상승 토픽의 트렌드이다. 상승 토픽은 정보보안, 고객만족, 개인정보, 서비스전략, 소통혁신으로 나타났다. <그림 6>은 연도별 하락 토픽의 트렌드이다. 하락 토픽으로

는 시스템구축 및 웹사이트로 확인되었다. <그림 7>은 연도별 중립 토픽의 트렌드이다. 중립 토픽은 기술수용모델, 혁신역량, 제품리뷰, 예측기술, 가상커뮤니티, e비즈니스신뢰, 서비스품질, 프로젝트리스크, 지식경영, 프로젝트교육, 아웃소싱역량, 공급망관리, 고객충성도, 서비스전략으로 나타났다.

표 4. 시계열회귀분석 결과

Table 4. Result of Time Series Regression (1990~2017)

Topic	regression coefficient	p-value	Hot/Cold	
T01	기술수용모델	-.339	.222	-
T02	혁신역량	-.107	.517	-
T03	제품리뷰	.086	.470	-
T04	예측기술	-.264	.263	-
T05	정보보안	.439	.000	Hot
T06	가상커뮤니티	.071	.660	-
T07	e비즈니스신뢰	-.336	.084	-
T08	서비스품질	-.025	.893	-
T09	프로젝트리스크	-.321	.121	-
T10	시스템구축	-1.136	.000	Cold
T11	지식경영	.021	.938	-
T12	개인정보	.339	.014	Hot
T13	웹사이트	-.436	.007	Cold
T14	프로젝트교육	-.343	.173	-
T15	아웃소싱역량	-.193	.265	-
T16	공급망관리	-.229	.148	-
T17	고객충성도	-.275	.182	-
T18	서비스전략	.175	.308	-
T19	고객만족	1.082	.001	Hot
T20	소통혁신	1.032	.001	Hot

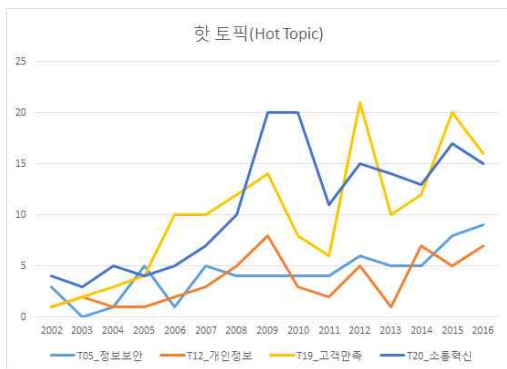


그림 5. 상승 토픽
Fig. 5. Hot Topic Trends

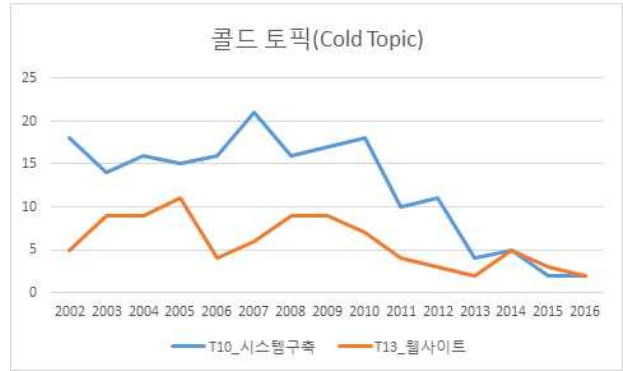


그림 6. 하락 토픽
Fig. 6. Cold Topic Trends

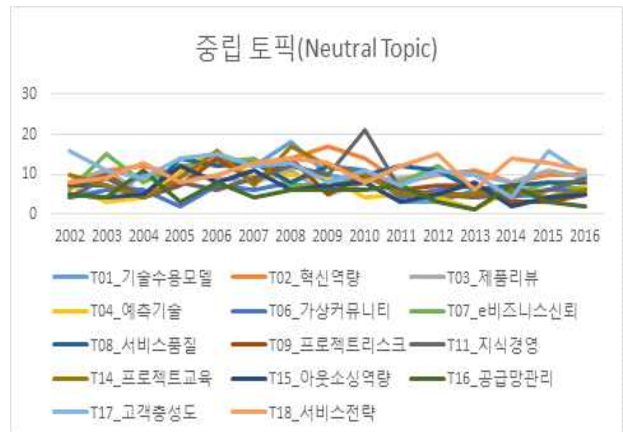


그림 7. 중립 토픽
Fig. 7. Neutral Topic Trends

4-3 저널별 토픽 동향분석

시간에 따른 저널별 핵심 토픽의 변화를 파악하기 위하여, 토픽모델링 결과 임계치 값(Cutoff value)을 저널별/연도별로 합산하였다. 저널별 토픽 동향분석 결과 2008년 214건을 정점으로 이전은 상승, 이후는 하락 추세를 보이고 있다. <그림 8>은 저널별 토픽 동향분석 결과이다.

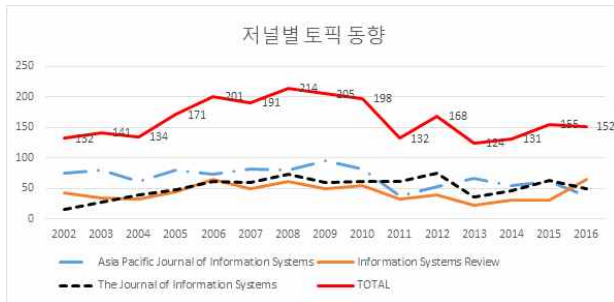


그림 8. 저널별 토픽 동향
Fig. 8. Topic Trends by Journal

V. 결론

본 연구에서는 2002년부터 2016년까지 15년간의 정보시스템 분야 국내선도 저널 3개에서 1,245개 논문의 초록을 대상으로 토픽모델링과 시계열회귀분석 방법을 사용하여, 연구동향을 분석을 수행하였다.

본 연구의 결과는 다음과 같다. 첫째, 토픽모델링결과 총 20개의 토픽(기술수용모델, 혁신역량, 제품리뷰, 예측기술, 정보보안, 가상커뮤니티, e비즈니스신뢰, 서비스품질, 프로젝트리스크, 시스템구축, 지식경영, 개인정보, 웹사이트, 프로젝트교육, 아웃소싱역량, 공급망관리, 고객충성도, 서비스전략, 고객만족, 소통혁신)이 도출되었다. 문서의 수 기준으로는 시스템구축, 혁신역량, 서비스전략, 고객충성도의 순으로 중요하게 나타났다. 둘째, 시계열회귀분석결과, 상승(Hot) 토픽으로는 4개(정보보안, 개인정보, 고객만족, 소통혁신)가 도출되었다. 하락(Cold) 토픽으로는 2개(시스템구축, 웹사이트)가 나타났다.

본 연구결과를 바탕으로 도출된 시사점은 다음과 같다. 첫째, 4차 산업혁명과 더불어 IT를 기반으로 한 혁신은 지속적으로 확대될 것이라고 판단되는 시점에 국내 정보시스템분야 연구동향 분석은 시의적절하다고 판단된다. 본 연구의 결과에서는 정보시스템분야에서 지속적으로 관심을 두어야 할 토픽으로 정보보호, 소통혁신과 고객만족이 도출되었다. 이 토픽들은 지속적으로 중요도가 높아지고 있다. 둘째, 기존의 많은 연구동향 분석에 있어서 활용되던 언어 네트워크분석기법의 한계점을 극복한 토픽모델링, 워드클라우드분석, 시계열회귀분석과 같은 연구기법을 적용하였고, 이는 연구동향을 더 깊이 있게 이해할 수 있게 한다. 특히, 시계열회귀분석을 활용한 상승/하락 토픽 분석은 연구동향분석에 통계적 유의성 검정을 추가할 수 있다는 점에서 더욱 더 의미가 있다.

본 연구는 여러 가지 시사점을 도출하였음에도 불구하고, 분석데이터의 다양성 관점의 한계가 존재한다. 보다 다양한 데이터를 기반으로 연구를 진행한다면 더 유의미한 시사점을 도출할 수 있을 것이다.

감사의 글

이 논문은 2015년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2015S1A3A2046711).

참고문헌

- [1] K. H. Lee, H. J. Jung and M. Song, "Weighted Subject - Method Network Analysis of Library and Information Science Studies," *Journal of The Korean Society for Library and Information Science*, Vol. 49, No. 3, pp. 457-488, 2015.
- [2] H. J. Kim, W. I. Park and C. S. Kim, "An Analysis of Trend of MICE Study using Language Network," *Journal of Tourism & leisure research*, Vol. 28, No. 3, pp. 199-217, 2016.
- [3] J. S. Park, C. S. Kim and K. Y. Kwahk, "Investigation of Research Trend in Hotel Domain Using Text Mining and Social Network Analysis," *Journal of Tourism & leisure research*, Vol. 28, No. 9, pp. 209-226, 2016.
- [4] D. M. Blei, "Probabilistic topic models," *Communications of the ACM*, Vol. 55, No.4, pp. 77-84. 2012.
- [5] M. Steyvers and T. Griffiths, Probabilistic topic models, in *Handbook of latent semantic analysis*, Vol. 427, No.7, pp. 424-440, 2007.
- [6] S. J. Yang, B. Y. Lee and H. W. Kim, "A Topic Modeling Approach to the Analysis of Happiness and Unhappiness," *Knowledge Management Research*, Vol.17, No.2, pp. 165-185, 2016.
- [7] J. S. Kim, E. J. Kwon and T. M. Song, "a Study on Using Social Big Data for Expanding Analytical Knowledge -Domestic Big Data supply-demand expectation," *Knowledge Management Research*, Vol. 15, No. 3, pp. 169-188, 2014.
- [8] K. R. Kim, H. J. Song and N.M. Moon, "Topic modeling for automatic classification of learner question and answer in teaching-learning support system," *Journal of Digital Contents Society*, Vol. 18, No. 2, pp.339-346, 2017.
- [9] R. Albright, *Taming Text with the SVD*. SAS Institute Inc., North Carolina: NC, Cary, 2004.

- 40, No. 5, pp. 1-23, November 2015.
- [10] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, "Indexing by latent semantic analysis," *Journal of the American society for information science*, Vol. 41, No. 6, pp. 391. 1990.
- [11] S. T. Dumais, G. W. Furnas, T. K. Landauer, S. Deerwester and R. Harshman, "Using latent semantic analysis to improve access to textual information," In *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM. pp. 281-285, May, 1988.
- [12] H. Y. Kwon, J. E. Kim and Y. T. Park, "Applying LSA text mining technique in envisioning social impacts of emerging technologies: The case of drone technology," *Technovation*, 2017.
- [13] G. E. Forsythe, C. B. Moler, and M. A. Malcolm, *Computer methods for mathematical computations*, Englewood Cliffs, NJ: Prentice-Hall, 1977.
- [14] N. Evangelopoulos, X. Zhang, and V. R. Prybutok, "Latent semantic analysis: five methodological recommendations," *European Journal of Information Systems*, Vol. 21, No.1, pp.70-86. 2012.
- [15] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, ch. 3, pp. 993-1022, January 2003.
- [16] J. S. Hyun, C. Y. Ko and K. E. Lee, "The Retrospect and Prospect on Korean Business Administration Literatures : Focused on KBR," *Korea Business Review*, Vol. 20, No.3, pp. 131-153, 2016.
- [17] I. S. Son and J. S. Kim, "Analysis of Trends in Research to Korean Strategic Alliance : Focused on the Korean and International Major Journals," *The Journal of Internet Electronic Commerce Resarch*, Vol. 15, No. 3, pp. 85-109, 2015.
- [18] W. G. Shim, "A Study on the Domestic Research Trend of Educational Service Quality," *Korean Review of Corporation Management*, Vol. 6, No. 1, pp. 55-76, 2015.
- [19] D. J. Kang and K. N. Lee, "A Study on Co-authorship Networks in Journal of Korea Trade Research Association using Social Network Analysis," *Korea trade review*, Vol. 40, No. 5, pp. 1-23, November 2015.
- [20] S. K. Hong, "Bibliometric Analysis on Research Trends of Service Management," *Journal of Korea Service Management Society*, Vol.17, No. 3, pp. 133-149, 2016.
- [21] S. J. Lee, "Research Trends of Korean Journalism and Communication Studies Using a Semantic Network Analysis," *Journal of the Korea Contents Association*, Vol.16, No.7, pp. 179-189, 2016.
- [22] S. D. Yang, "Semantic Network Analysis of Research Trend Related to Private Security," *Journal of the Korea Contents Association*, Vol.13, No. 11, pp. 894-901, 2013.
- [23] Y. B. Jung and E. S. Park, "Keyword Analysis of Two SCI Journals on Rock Engineering by using Text Mining," *Tunnel & Underground Space*, Vol. 25, No. 4, pp. 303-319, 2015.
- [24] J. Y. Lee, J. Y. Moon and H. J. Kim, "Examining the Intellectual Structure of Records Management & Archival Science in Korea with Text Mining," *Journal of The Korean Society for Library and Information Science*, Vol. 41, No. 1, pp. 345-372, 2007.
- [25] S. H. Seo and H. Y. Lee, "Fintech trend analysis using topic modeling of BM patents," *The Korean Institute of Industrial Engineers*, Seoul, pp. 471-480, 2015.
- [26] H. C. Kang, Data mining methodology for big data analysis, *Free academy*, 2014.
- [27] K. H. Lee and K. Y. Kwahk, "Effects of Word-of-Mouth and Assurance on Trust in the Internet Shopping Mall Environments: The Moderation Effect of Ease of Product Evaluation," *Knowledge Management Research*, Vol. 15, No. 3, pp. 141-168, 2014.
- [28] S. H. Jeon and K. Y. Kwahk, "Effects of Justice and Authenticity on Recovery Satisfaction in the Electronic Commerce Environments: The Moderation Effects of Involvement," *Knowledge Management Research*, Vol. 16, No. 1, pp.71-93, 2015.
- [29] T. L. Griffiths and M. Steyvers, "Finding scientific topics," in *Proceedings of the National academy of Sciences*, Vol. 101, No. 2, pp. 5228-5235, April 2004.



김 창 식 (Chang-Sik Kim)

2002년 : 경희대학교 산업정보대학원 (경영학석사)
2013년 : 국민대학교 비즈니스IT전문대학원 (경영정보학박사-비즈니스IT)

1995년~2014년: 한화호텔앤드리조트/한화에스앤씨

2015년~현재: 국민대학교 비즈니스IT전문대학원 BK21 플러스 사업팀 계약교수

※ 관심분야 : 텍스트마이닝(Text Mining), 머신러닝(Machine Learning), 딥러닝(Deep Learning),
데이터 애널리틱스(Data Analytics), 환대관광정보(Hospitality and Tourism Information) 등



최 수 정 (Su-Jung Choi)

2016년 ~ 현재: 국민대학교 비즈니스IT전문대학원 석사과정

※ 관심분야 : 소셜네트워크분석 및 응용(Social Network Analysis and Its Application),
데이터 애널리틱스(Data Analytics) 등



곽 기 영 (Kee-Young Kwahk)

1988년 : 서울대학교 경영대학 (경영학사)
1990년 : 한국과학기술원 경영학과 (경영학석사)
1999년 : 한국과학기술원 테크노경영대학원 (경영정보학박사)

1990년~2002년: 삼성전자/삼성SDS

2003년~2005년: 계명대학교 경영대학

2006년~현재: 국민대학교 경영대학/비즈니스IT전문대학원 교수

※ 관심분야 : 소셜네트워크분석 및 응용(Social Network Analysis and Its Application),
데이터 애널리틱스(Data Analytics), 지식경영(Knowledge Management) 등