

# 텍스트 마이닝 기법을 이용한 정보시스템 분야 연구 동향 분석

## Exploring Dynamics of Information Systems Research Trend Using Text Mining Approach

안 정 국 (Jungkook An) 연세대학교 정보대학원 박사과정  
김 소 담 (Sodam Kim) 연세대학교 정보대학원 박사과정  
김 희 웅 (Hee-Woong Kim) 연세대학교 정보대학원 교수, 교신저자

### 요약

최근 정보통신기술(Information and Communication Technology) 및 사물인터넷 시대가 도래함에 따라 융복합 환경에 따른 다양한 기술의 발전이 이루어지고 있다. 이에 따라 관련 학문에 대한 이론 및 활용 기술에 대한 관심이 고조되고 있는 상황이다. 이러한 패러다임의 변화는 학문들 간의 급격한 융복합 현상을 초래하였으며, 특히 정보시스템학(Information Systems)은 이러한 변화를 선도해 왔다. 정보시스템학은 다른 학문들과의 관계에 있어서 분화적(Divergence) 역할에서 나아가 융합적(Convergence) 역할까지 수행하고 있으나, 이러한 연구 동향에 관한 연구는 부족한 실정이다. 따라서, 본 연구에서는 시간 경과에 따른 정보시스템의 연구동향을 비교 분석하여 핵심 개념들을 살펴봄으로써 향후 정보 시스템학 연구의 방향에 대한 시사점을 찾고자 한다. 구체적으로, 1980년부터 2015년까지의 경영 정보학의 상위 국제저널 48,102개의 논문제목, 저자, 초록, 키워드 분석을 통해 저자들의 공동 연구 네트워크 분석 및 연구 토픽 추출 결과를 연대별로 비교 분석하여 시각화하였다. 본 연구의 결과가 정보시스템 분야의 연구자들에게 정보시스템의 정체성에 대한 폭넓은 이해와 향후 연구 방향에 대한 새로운 시사점을 주기를 기대한다.

**키워드 :** 정보시스템, 연구동향, 공동연구 네트워크, 토픽모델링, 네트워크 분석

## I. 서론

최근 인터넷과 ICT(Information and Communication Technology)의 발달로 다양한 정보와 기술의 융복합적 활용이 확산되고 있다. 기술의 발전으로 인해 끊임없이 생산되는 데이터를 어떻게 유용한 정보로 활용할 수 있는가에 대한 질문은,

분야를 넘어서 현대사회의 최대 이슈라고 할 수 있다. 또한, 결국 이에 대한 답을 다시 기술로부터 찾아야 한다는 점에서, 정보와 기술의 융복합적 활용의 필요성 및 확산 양상의 당위성을 찾을 수 있다. 나아가 이러한 융복합의 패러다임이 다양한 학문적, 그리고 실무적 분야에서 꾸준히 일어나고 있다는 측면에서 그 규모 및 영향력도 상

당하다. 특히, 정보시스템(Information System, IS)은 학술적 연구에 있어서 다양한 학문들의 실제적 활용 및 적용이라는 학문적 융복합의 패러다임의 중요한 역할을 담당하고 있다. 기존 IS 분야에 대한 연구를 살펴보면, 순수정보시스템(Pure IS)과 융합정보시스템(Hybrid IS)으로 나눌 수 있다(Walstrom and Hardgrave, 2001). Walstrom과 Hardgrave는 연구자 그룹을 대상으로 설문 조사를 하여, 주요 IS 저널에 게재된 논문을 순수정보시스템, 융합정보시스템 그리고 비정보시스템 분야로 구분하였다. 구체적으로 순수정보시스템에 속한 저널은 J8에 따른 인용 값(citation value)이 8.8% 이상이 되는 저널을 말하며, Information & Management, Journal of Global Information Management, Information Systems Management, Journal of Computer Information Systems, and DATA BASE for Advances in Information Systems 등의 13개의 저널이 속해있다. 또한, 융합정보시스템은 인용 값의 중앙값(median)인 0.8% 이상이 되는 저널을 말하며, 이에 속한 저널은 Journal of Systems and Software, International Journal of Human-Computer Studies, and Decision Sciences, Decision Support Systems 등이 있다. Chan *et al.*(2015)은 이와 같은 분류를 기반으로 비정보 시스템 분야를 제외한 두 개의 분류로서 설문조사 및 네트워크 분석을 통해 저널 간의 관계에 대해 분석한 바 있다. 이에, 본 연구에서는 1980년부터 2015년까지의 정보시스템 분야 논문 48,102개를 과학 인용 온라인 데이터베이스인 Web of Science에서 추출하여, 정보시스템 저널의 가중치와 인용 수에 따라 순수정보시스템과 융합정보시스템으로 분류한 방법을 사용하여 저널들을 분석하여 기존 연구의 확장을 시도하였다(Chan *et al.*, 2006; Chan *et al.*, 2015). 구체적으로, 추출 및 분류한 논문들의 제목, 키워드, 그리고 초록을 중심으로 주요 어휘에 대해 토픽모델링(Topic Modeling) 및 빈도 분석(Frequency Analysis), 네트워크 분석(Network Analysis)을 수행함으로써 순수정보시스템과 융합정보시스템 연구 주제의 핵심적 개념이

무엇인지 살펴보고자 하였다.

따라서 본 연구는 (1) 정보시스템 연구 동향의 시간에 따라 어떠한 핵심적 주제들이 있는지, (2) 순수정보시스템과 융합정보시스템의 시간의 흐름에 따른 주제들 중에 핵심키워드는 어떠한 것이 있는지, 그리고 (3) 정보시스템 연구의 허브 역할을 하는 연구자는 누구인지에 대하여 알아보게 하였다. 이를 기반으로 정보시스템 연구 동향을 심화적으로 파악 및 분석하여 향후 정보시스템학 연구 발전을 위한 기초자료와 통찰력을 제공할 수 있기를 기대한다. 또한 시간적 경과 추이에 따른 순수정보시스템과 융합정보시스템의 핵심적 개념을 분석하여, 정보시스템 학문이 다른 학문 및 실무적 관계에서 융합적인(Convergence) 역할에서 나아가 분화적인(Divergence) 역할까지 수행하고 있는지에 대한 검증도 시도하였다. 마지막으로, 연구의 허브역할을 하는 연구자들을 공동저자 네트워크로 분석함으로써 정보시스템 연구에 있어 중추적인 역할을 하는 연구자들을 나타내고 정보시스템 연구의 공동저자 네트워크를 제안하고자 한다.

본 연구의 구성은 다음과 같다. 제II장에서는 연구의 결과 도출을 위해 활용된 토픽모델링 방법론과, 융합과학(Convergence Science) 및 분화과학(Divergence Science)의 개념적 배경에 대해 살펴보고, 제III장에서는 본 연구의 설계 방법론에 대해 논의한다. 그리고 제IV장의 연구 분석 결과에 이어, 마지막 제V장에서는 본 연구의 결론과 시사점을 제시하고자 한다.

## II. 개념적 배경

### 2.1 토픽모델링(Topic Modeling)

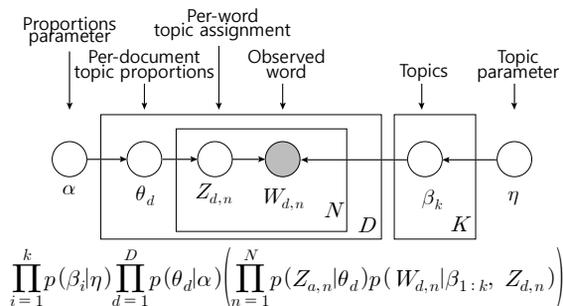
토픽모델링은 각 문헌을 주제(Topic)의 확률적 혼합체로 표현하고, 텍스트 정보에서 숨겨져 있는 주제들을 찾기 위해 고안된 통계적 추론 기법이다(Blei, 2012; Steyvers and Griffiths, 2007; 강범일

등, 2013). 가장 대표적으로 활용되고 있는 LDA (Latent Dirichlet allocation) 알고리즘은 높은 성능과 편의성으로 인해 토픽모델링 분야에서 표준적 방법론으로 인식되고 있다(Blei *et al.*, 2003). LDA의 대략적인 분석방법은 다음과 같다. 각 문서의 주제는 소수의 단어 혼합물로 구성되어 있으며, 각 단어의 출현은 문서의 주제에 의해 결정된다고 가정하여 토픽모델링이 수행된다(Blei *et al.*, 2010). 즉, 특정 주제들의 집합이라고 가정된 한 문헌을 구성하는 단어들을 확률적으로 계산하여, 이 결과값을 토픽 주제어들의 집합으로 추출하는 알고리즘이다. 토픽모델링에서는 단어가 서로 독립적이지 않다는 가정(Dirichlet Distribution)에 기초하여 단어 생성 조건에 따라 사후 확률을 추론하며, 이는 <그림 1>과 같이 Blei(2012)의 확률적 그래프 모델로 표현된다.

LDA는 알고리즘 자체의 단순성을 비롯해, 의미적으로 일관성이 있는 주제를 생산할 수 있다는 점 등에서 다양한 장점을 가진다(Mimno *et al.*, 2008). 방대한 데이터의 처리 및 분석이 요구되면서, 소셜 미디어(Social Media), 신문기사 등 텍스트 기반의 비정형 데이터에 대한 연구는 감성 분석과 더불어 LDA를 통해 점진적으로 이루어지고 있다(안정국, 김희웅, 2015; 조승연 등, 2015). 트위터(Twitter) 상에서 논의되는 토픽 분포와, 같은 기간 내 뉴욕 타임지(New York Times) 기사의 주

제 분포를 비교한 연구 등의 소셜 네트워크 서비스(SNS)를 통한 토픽모델링의 연구도 있었다(Zhao *et al.*, 2011; 강성경 등, 2015). 또한, 신문 그 자체로도 시사적이며 시의적 주제를 지녔다는 점과 그 파급력의 특성을 기반으로 토픽모델링을 활용하여 여러 차례 분석이 진행되어 왔다(Grimmer, 2010; Kam and Song, 2012; Newman and Block, 2006).

이러한 토픽모델링을 활용하여, 생물정보학이나 문헌정보학 등과 같이 다양한 학계에서 국내 및 국외의 학술자료 연구 동향을 파악해, 연구자에게 의미 있는 가이드를 제공하고자 하는 연구도 진행되어 왔다. Gerrish and Blei(2010)는 시간적 흐름에 따라, 논문 코퍼스(Corpus)로부터 주제 변화를 파악 후, 개별 문헌의 파급력 및 영향력을 분석한 바 있다. 또한, 생물정보학 분야의 논문을 연구 대상으로 하여, 화두가 되고 있는 연구 주제들을 추출, 연구 경향을 살펴보고자 한 연구도 있었으며, 국내에서는 문헌정보학 논문을 대상으로 진행된 연구도 있다(Song *et al.*, 2013; 박자현, 송민, 2013). 하지만, 타 학문에 대해 융복합적 특징을 가졌으며, 사회 전반에 영향력을 미치고 있는 정보시스템학 분야에서는 아직까지 어떠한 연구가 어떻게 진행되어 왔는지에 대한 고찰이 부족한 실정이다(Chen and Hirschheim, 2004). 토픽모델링은 아래 <그림 1>과 같은 LDA 모형에 따라, 각종



Note:  $\alpha$  = Proportions parameter,  $\theta_d$  = Per-document topic proportions,  $Z_{d,n}$  = Per-word topic assignment,  $W_{d,n}$  = Observed word,  $\beta_k$  = Topics,  $\eta$  = Topic parameter

<그림 1> LDA 확률 그래프 모델(Blei, 2012)

문서를 다양한 주제들로 연관 추론할 수 있는 장점이 있다(Blei *et al.*, 2003; Blei, 2012; 박자현, 송민, 2013). 이를 활용하여 생물 정보학, 문헌정보학 등 다양한 분야의 학문에서 “저널(Journal)”이라는 전문적 매체를 이용해 학문적 경향을 파악하고자 하는 다수의 연구가 진행되고 있다(Gerrish and Blei, 2010; Song and Kim, 2013; Song *et al.*, 2014; 박자현, 송민, 2013; 이석, 정병규, 2015).

## 2.2 융합과학(Convergence Science)

융합과학은 과학, 기술 및 인문사회과학 등 세분화된 학문들의 복합적 결합, 통합 및 응용에 의해 만들어진 새로운 과학분야라고 할 수 있다. 20세기 중엽부터 21세기 즉, 근대에서 현대로 넘어가는 시기에 이르러 본격적으로 학문과 기술의 수렴 및 융합의 흐름이 전개되었다(Négyessy *et al.*, 2008). 각각 다른 성격과 특성을 내포하고 있다고 여겨졌던 다양한 학문들이 서로 물리적, 추상적으로 공통적 법칙들을 가지고 있다는 사실이 증명되었고, 이는 학제 간 다양하고 활발한 연계적 연구를 이끌었다. 융합과학의 대표적 사례는 세계적 학술 연구기관인 “MIT 미디어랩”을 들 수 있다. 이 기관은 공학에 예술과 인문학 등 다소 이질적인 학문을 새롭게 접목시키며 ‘상상력을 발전시킨다’라는 슬로건을 가지고 다양한 방향으로 연구를 해왔다. MIT 미디어랩은 이러한 다각적 연구를 통해 세계 모든 어린이를 위한 100달러 미만의 컴퓨터, 스스로 생각하는 인공지능(Artificial Intelligence, AI) 로봇, 옷처럼 입을 수 있는 컴퓨터 등 창의적이고 생산적인 기술을 만들어 인류에 혁신을 가져다 주고 있다. 이와 같이 과학의 발달은 새로운 국면으로 접어들고 있다. 융합과학은 인문학, 사회과학, 예술, 공학, 과학 및 문화 등 여러 영역의 학문들을 창조와 융합의 원리로 탐구하여 인간의 삶뿐만 아니라 인간성을 향상하는 것이 융합과학이 궁극적으로 추구하고자 하는 목적이다. 또한, 자연이라는 하나의 대상

을 각 학문의 개별적 특성은 유지하되 각각의 학문적 특성을 고려하여 통합적인 탐구를 함으로써 유의미한 종합적 시사점을 이끌어내는 것이 융합과학의 목표라고 할 수 있다.

## 2.3 분화과학(Divergence Science)

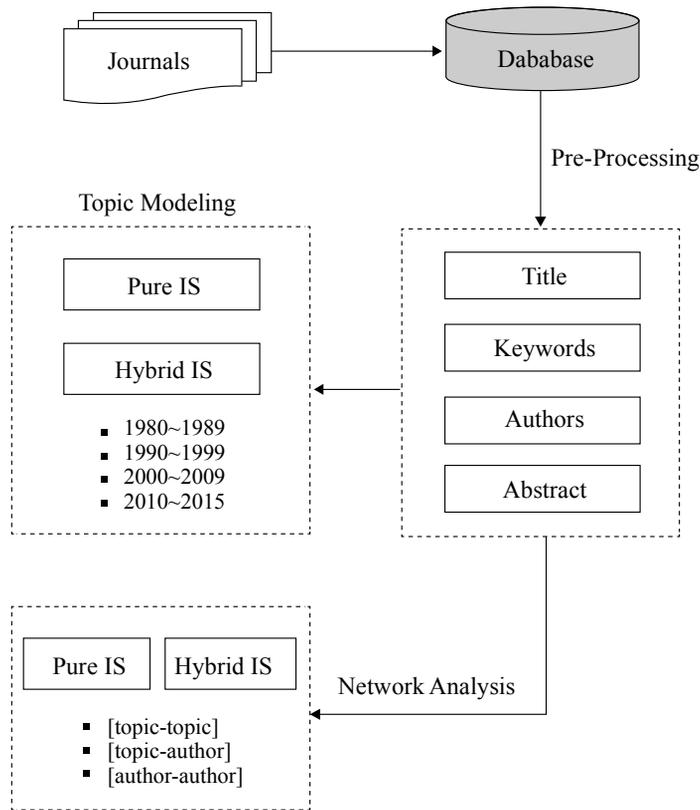
과거 본래의 학문은 나뉘어져 있지 않았고 하나의 뿌리를 가지고 있었다. 하지만, 근래 18세기 말부터 산업혁명을 거치며 연구의 깊이가 더해지고, 인류가 세상을 변화시키기 위해 학문 체계가 분화(Divergence)되어 더욱 세밀히 연구되어 왔다. 이러한 의미에서 분화과학이란, 공통적인 학문이라는 토대 안에 있지만, 크게 자연과학과 사회과학 등으로 나뉘어 각각 상이한 연구대상을 탐구하고, 예측(Prediction)과 의사소통(Communication) 측면, 법칙(Law)과 관습(Convention)적 측면, 연역적 측면과 귀납적 측면에서 연구를 한다는 점에서 아래의 <표 1>과 같이 차이를 가지고 구분되어 연구되어왔다(Fisher, 2001; Losee, 1972).

<표 1> 사회과학과 자연과학

	사회과학	자연과학
연구대상	인간 혹은 그들의 의도적 행위	객관의 세계
설명력	의사소통 기반	예측 기반
일반화 정도	관습의 속성	법칙의 속성
추론법	귀납법	연역법

## III. 연구 설계

본 장에서는 연구에 관련된 데이터 수집 과정과 분석 프로세스(토크포모델링, 네트워크 분석)에 대해 설명하고자 한다. 대략적인 데이터 분석 프로세스는 <그림 2>와 같이 도식화된 형태로 나타나 있다. 이러한 과정을 토대로 연구를 진행하여 결과를 분석한 후, 시사점을 도출하고자 한다.



〈그림 2〉 텍스트 마이닝 프로세스

### 3.1 데이터 수집

본 연구는 국제 논문(SCIE, SSCI, A&HCI)을 실시간 검색 가능할 수 있도록 하는 웹 데이터베이스인 “Web of Science”에서 1980년부터 2015년까지 연구된, 약 35년여 간의 정보시스템(IS) 분야 논문 데이터를 수집하였다. 수집된 정보시스템 분야 전체 논문 수는 총 48,102개였으며, 그 중 순수정보시스템 관련 논문은 10,876개 그리고 융합정보시스템 관련 논문은 37,226개였다. 기존 연구(Chan *et al.*, 2006)에 기반하여 저널들을 분류 및 수집하였으며, 분류한 저널 카테고리는 <표 2>와 같다.

데이터 수집 절차를 살펴보면, 저널 카테고리

에 따라서 저널 명, 논문 제목, 초록, 저자, 키워드, 추가 키워드, 발행 연도, 카테고리, 인용 수, 인용 번호 등을 수집하여 데이터를 구성하였다. 각 연도 구간별 데이터 수집 결과는 아래의 <표 3>과 같다. 또한, 웹 사이트를 통해 수집된 데이터에는 각종 태그와 마크 업 기호 등이 존재하였으므로, 총 48,102건의 데이터는 전처리(preprocessing)하여 결과를 도출해 내고자 하였다. 구체적으로, 데이터 분석 시 본 연구와 크게 관련이 없다고 여겨지는 형용사, 조사, 부사 등을 제거하는 등의 정제 과정을 거쳤으며, 문장 경계 인식, 자동 띄어쓰기 등 형태소 분석 단계 이전에 필요한 전처리 과정을 진행하였다(최기선 등, 2012).

〈표 2〉 정보시스템분야 저널 분류(Chan *et al.*, 2006)

Classification	Journals
Pure IS	<ul style="list-style-type: none"> <li>- MIS Quarterly</li> <li>- Information Systems Research</li> <li>- Journal of Computer Information Systems</li> <li>- Information Systems Management</li> <li>- Journal of Global Information Management</li> <li>- Journal of Management Information Systems</li> <li>- DATA BASE for Advances in Information Systems</li> <li>- Journal of Information Technology</li> <li>- Information Systems Journal</li> <li>- Journal of Strategic Information Systems</li> <li>- Information Systems Journal</li> <li>- European Journal of Information Systems</li> <li>- Information &amp; Management</li> <li>- Journal of the Association for Information Systems</li> </ul>
Hybrid IS	<ul style="list-style-type: none"> <li>- Expert Systems with Applications</li> <li>- ACM Transactions on Database Systems</li> <li>- International Journal of Human Computer Studies</li> <li>- The Journal of Information Systems</li> <li>- Human Computer Interaction</li> <li>- Journal of Systems and Software</li> <li>- Knowledge based Systems</li> <li>- Behaviour Information Technology</li> <li>- IEEE Transactions on Software Engineering</li> <li>- Computers in Human Behavior</li> <li>- Informs Journal on Computing</li> <li>- IEEE Transactions on Knowledge and Data Engineering</li> <li>- Datamation</li> <li>- ACM Computing Surveys</li> </ul>

〈표 3〉 정보시스템분야 논문 데이터 수집 결과

Classification	Year	Number of papers
Pure IS	1980~1989	476
	1990~1999	2,246
	2000~2009	4,619
	2010~2015	3,535
Hybrid IS	1980~1989	4,832
	1990~1999	8,147
	2000~2009	9,741
	2010~2015	14,506
Total	1980~2015	48,102

### 3.2 데이터 분석

#### 3.2.1 토픽모델링 분석

현대 사회에서 정보량은 기하급수적으로 증가하는 추세이다. 이에 원하는 정보를 단번에 찾는

것이 점점 어려워지고 있으며, 심지어 대량의 데이터 내에서 주어진 데이터가 어떤 정보 및 가치를 포함하고 있는지 파악하지 못할 때도 있다. 따라서, 수많은 데이터를 탐색해 정보 추출을 도와주고, 숨겨진 패턴을 정리해주는 분석법이 필요하기에 이를 해결해줄 수 있는 토픽모델링이 급속도로 부상하였다. 토픽모델링이란 수많은 문헌 모델링 기법 중 대표적 분석법을 말한다. 특히, Blei *et al.*(2003)가 제안한 초창기 토픽모델링 기법 중 하나인 LDA(Latent Dirichlet allocation) 알고리즘이 다양한 분야에서 활발히 적용되고 있는 추세이다. LDA 알고리즘은 문헌을 기반으로 하는 절차적 확률분포 모델로써 문헌은 특정 확률에 의해 선택된 단어들로 구성된 토픽들의 집합으로 표현된다(Blei *et al.*, 2003; Blei, 2012). 본 연구는 수집된 데이터에 들어있는 저널 명(Journal), 논문 제목

(Title), 초록(abstract), 저자(author 1-10) 등의 정보를 토픽모델링 LDA 알고리즘으로 분석해 그에 대한 본 연구의 함의를 이끌어내고자 하였다.

### 3.2.2 네트워크 분석

토픽모델링 수행 후 분석의 두 번째 단계로써 네트워크 분석을 수행하고자 한다. 네트워크 분석은 1930년대부터 인류학, 사회 심리학, 사회학 등 사회과학 분야에서 활발한 연구가 진행된 분석법이다(정근하, 2011). 현재는 자연과학, 융합과학, 경영학 등 많은 분야에서 본 이론이 응용되고 있다. 그 중 사회 네트워크 분석(Social Network Analysis, SNA)이 사회과학 분야에서 가장 많이 응용되어 왔다고 할 수 있다(Bonacich, 1987; Galaskiewicz and Stanley, 1993). 사회 네트워크 분석의 장점은 복잡한 관계 내에서 구조적 유사성을 탐색적으로 찾아낼 수 있다는 것이다. 또한, 분석 대상의 여러 가지

속성(예: 성별, 직위, 소속 등)에 의해 개별 참여자들이 가지고 있는 구조적 특징을 통계 분석하여 설명할 수 있다는 점에서도 매우 유용하다. 본 연구에서는 위에서 언급된 네트워크 분석의 장점을 통해 정보시스템 학문 분야의 시대별 이슈가 되는 연구 키워드(Research Keywords)와 그 시대 연구를 선도하고 있는 연구자(author) 속성을 활용하여 공동저자 네트워크 분석을 시행하였다.

## IV. 연구 결과

### 4.1 토픽모델링 기반의 정보시스템 학문 동향

#### 4.1.1 정보시스템 토픽(1980~1989)

본 연구에서는 1980년부터 1989년 약 10년 간 순수정보시스템(Pure IS) 분야에서 연구된 논문 총 476개를 수집할 수 있었다. 위 데이터들을 토

<표 4> 순수정보시스템 토픽(1980~1989)

Topic	Description	
	Label	Terms
01	시스템 분석	systems, information, database, design, system, databases, relational, model, management, approach, mis, analysis, distributed, development, decision, computer, based, support, retrieval, structure
02	효과	framework, study, human, comparing, communication, business, effectiveness, field, planning, effects, development, statistical, semantics, office, incomplete, values, tree, queries, variable, limitations
03	조직	language, organizational, planning, natural, network, implementation, managers, organization, fuzzy, context, untitled, impact, constrained, program, trends, change, research, integrity, dependencies
04	요인	data, base, bases, requirements, types, storage, environment, trees, modeling, analysis, managing, logical, dependency, factors, organizational, resource, issues, issue, automatic, attribute
05	상호작용	untitled, utilization, key, success, multiple, office, extending, selection, design, algorithmic, fortran, interaction, rdbm, turkish, definitions, library, searching, corporations
06	인터페이스	level, multi, word, dependency, factors, superimposed, 1st, special, semantic, attribute, application, algebra, integration, interface, modeling, algorithm, donau, acceptance, scheduling, uniform
07	컴퓨팅	coding, processing, system, files, tool, incorporating, searching, software, dictionary, index, technique, definition, computing, support, dependencies, semantics, semantic, knowledge, heterogeneous, model
08	평가	access, control, concurrency, characteristics, resource, problems, graphical, interfaces, implications, mis, secondary, clustering, views, evaluation, issues, schemas, context, intelligent, queries, relation
09	효율	problem, issues, file, relationship, allocation, function, effectiveness, empirical, project, calculus, versus, decision, work, mechanisms, application, logic, oriented, object, objects, information
10	기술	users, functional, search, trees, chaining, systematic, programming, time, technology, mis, pattern, relations, formal, structured, conceptual, linear, cost, databases, design, efficiency

대로 LDA 알고리즘을 통해 아래와 같은 결과를 도출할 수 있었으며, <표 4>는 각 주제에 해당하는 순수 정보시스템 중 논문 투고 빈도가 높은 핵심 연구 키워드 상위 20개를 요약한 결과를 나타낸다. 앞서 언급한 바와 같이, LDA 모델은 텍스트 형식 데이터의 단어들과 구조를 통해 문서를 테마로 분류하는데, 각 주제의 제목은 연구자에 의해 가장 빈번하게 쓰이는 단어들로 추론하여 판단할 수 있다(박자현, 송민, 2013). 본 연구는 시스템 분석, 효과, 조직, 요인, 상호작용, 인터페이스, 컴퓨팅, 평가, 효율 그리고 기술의 10가지 주제가 도출된 LDA 결과를 이용하였다. 이는

1980~1990년의 순수정보시스템 핵심 연구로 시스템 분석과 조직 평가 관련 연구가 성행했음을 알 수 있다.

1980년부터 1989년 약 10년 간 융합정보시스템 (Hybrid IS) 분야에서 연구된 논문 총 4,832개를 수집할 수 있었다. <표 5>에서 확인할 수 있듯이, 1980~1990년의 융합정보시스템 핵심 연구에 대한 LDA 토픽모델링 결과는 데이터/IBM/프로토콜, 네트워크/이해, 제어/적용/생산성, 서베이/방법론/지식, 통합/도구/인텔리전스, 어플리케이션, 시간/포트란/성능, 상관관계/알고리즘, 자동화/산업, 유닉스/운영과 같은 10가지의 정보시스템 기

<표 5> 융합정보시스템 토픽(1980~1989)

Topic	Description	
	Label	Terms
01	데이터, IBM, 프로토콜	data, distributed, processing, query, states, databases, united, fault, news, tolerant, protocol, structures, abstract, managing, quality, training, ibm, whats, micro, micros
02	네트워크 이해	network, users, local, dp, concurrent, ada, partial, specifications, distributed, form, understanding, key, programming, operations, prototyping, nets, petri, expert, ibm, product
03	제어, 적용, 생산성	control, language, user, process, concurrency, access, oriented, environment, efficient, object, standards, dynamic, metrics, complexity, implementation, directed, driven, productivity, net, natural
04	서베이, 방법론, 지식	based, model, support, survey, distributed, machines, knowledge, blue, interactive, methodology, strategy, automatic, ai, entity, decision, logic, relationship, dynamic, state, verification
05	통합, 도구, 인텔리전스	languages, programming, storage, survey, pc, lan, intelligence, implementation, market, change, integration, testing, artificial, tool, power, business, semantic, phase
06	어플리케이션	specification, life, human, engineering, methods, future, cycle, files, micro, large, open, factors, dp, errors, test, trees, hashing, error, applications, faults
07	시간, 포트란, 성능	time, high, real, level, top, cost, back, end, tech, hard, issue, behavior, waiting, change, engineering, performance, discrete, issues, pascal, fortran
08	상관관계, 알고리즘	software, system, design, analysis, development, distributed, performance, program, relational, language, reliability, model, engineering, testing, correction, models, programming, big, databases, algorithm
09	자동화, 산업	applications, technology, computers, protocols, special, approach, computing, industry, road, method, automation, ibm, transaction, detection, ada, managers, representation, research, recovery
10	유닉스, 운영	systems, database, computer, information, management, networks, programs, evaluation, approach, algorithms, ibm, unix, file, system, modeling, communication, study, resource, queuing, operating

술에 관련된 주제로 도출되었다.

4.1.2 정보시스템 토픽(1990~1999)

논문 2,246개를 기반으로 한 1990~1999년의 순수정보시스템 핵심 연구에 대한 LDA 결과는 네트워크, 통합, 전략, 엔지니어링, 웹, 시스템, 생산성, 네트워크, 상호작용 그리고 분석의 10가지 주제로 도출되었다. <표 6>에서 볼 수 있듯이, 토픽 모델링 결과는 네트워크 분석을 통해 기존의 순

수정보시스템 연구 주제는 기존의 시스템들의 통합이라는 연구가 성행했음을 알 수 있다.

<표 7>은 8,147개를 기반으로 하여, 1990년부터 1999년까지 이루어진 융합정보시스템 연구의 LDA 토픽모델링 결과를 나타낸다. 10가지 주제는 프로그램/알고리즘, 성능, 로직, 자동화, 네트워크, 처리/비용/효율, 서비스/저장으로 나타났으며, 융합정보시스템 연구에서 네트워크의 성능에 관한 연구가 활발히 이루어 졌음을 알 수 있다.

<표 6> 순수정보시스템 토픽(1990~1999)

Topic	Description	
	Label	Terms
01	네트워크	design, approach, environment, business, neural, strategic, network, planning, research, programming, modeling, methodology, evaluation, outsourcing, implications, linear, distributed, professionals, managing, interactive
02	통합	technology, learning, special, dss, electronic, performance, implementation, modeling, integrating, research, work, issue, development, formal, task, building, structured, tools, implementing, integration
03	전략	business, mis, strategy, language, control, intelligent, reengineering, optimization, software, group, access, industry, strategic, fuzzy, level, electronic, integrated, introduction, tools, reasoning
04	엔지니어링	study, case, empirical, relational, databases, analysis, investigation, evaluation, engineering, perspective, software, organizations, modeling, project, generation, implementation, power, automatic, dynamic, field
05	웹	based, model, knowledge, case, internet, web, software, applications, logic, rule, skills, introduction, conceptual, spatial, multiple, constraint, context, practice, department, decisions
06	시스템	systems, information, system, management, database, development, retrieval, analysis, distributed, time, expert, approach, comparison, query, quality, application, managing, data, role, processing
07	생산성	object, oriented, theory, planning, computing, untitled, student, economic, success, process, framework, productivity, transaction, teaching, methodology, tool, issue, operation, inference, function
08	네트워크	data, computer, user, networks, end, issues, interface, problems, base, users, human, communication, service, databases, requirements, text, supported, schema, demand, formulation
09	상호작용	organizational, group, framework, strategies, server, client, success, change, issue, problem, conceptual, queries, multimedia, effects, large, integration, inter, evaluating, solving, interactive
10	분석	decision, support, process, making, models, analysis, relationship, role, research, commerce, factors, method, processes, entity, markets, memory, semantics, service, face, supporting

<표 7> 융합정보시스템 토픽(1990~1999)

Topic	Description	
	Label	Terms
01	프로그램, 알고리즘	programming, languages, representation, fault, evaluation, neural, network, security, complexity, search, algebra, algorithms, databases, trees, constraints, tolerance, nets, learning, protocols, diagnosis
02	성능	time, real, performance, level, transaction, safety, formal, high, requirements, probability, scheme, critical, symbolic, resolution, net, conceptual, system, world, meta, cognitive
03	로직	databases, control, algorithms, concurrency, engineering, logic, models, scheduling, verification, parallel, algorithm, model, discovery, computing, reasoning, classification, method, functional, semantics, relational
04	자동화	software, design, reliability, analysis, relational, distributed, development, learning, semantic, query, quality, error, consistency, detection, maintenance, environment, automatic, algorithm, mining, requirements
05	네트워크	computer, networks, programs, methods, memory, problem, temporal, distributed, machine, logic, multimedia, solving, generation, human, models, interface, processing, formal, databases, testing
06	처리, 비용, 효율	performance, language, query, processing, program, simulation, theory, queries, estimation, technology, architecture, reuse, nets, indexing, acquisition, concurrent, cost, parallel, measurement, rules
07	서비스, 저장	systems, system, database, information, management, expert, retrieval, artificial, intelligence, recognition, pattern, multiple, storage, complex, distributed, measurement, function, file, service
08	지식, 방법론	knowledge, based, object, oriented, rule, case, model, constraint, base, domain, acquisition, content, orientation, methodology, component, representation, web, explanation, scenario, tree
09	분석, 적용	data, analysis, decision, support, flow, reasoning, deductive, distributed, state, logic, graph, active, failure, architecture, hypermedia, formal, load, implementation, abstract, query
10	사양, 최적화	model, specification, process, optimization, user, modeling, petri, models, testing, metrics, integration, access, temporal, interface, checking, intelligent, locking, semantics, verification, communication

4.1.3 정보시스템 토픽(2000~2009)

2000~2009년의 순수정보시스템 핵심 연구의 토픽모델링 LDA 결과는 논문 4,619개를 분석하여 도출되었다. <표 8>에서 확인할 수 있듯이, 정보 기술, 네트워크 정보, 인플루언스, 기술, 전략, 공유, 시장, 관계 그리고 지식이라는 주제로 도출되었다. 이 시기는 모바일 시장의 성장으로 정보의 생성과 수집이 용이해진다는 점에서 특징을 갖는다. 따라서, 순수정보시스템에서의 정보의 과급성에 대한 연구의 관심은 정보의 공유 관계, 기술, 영향력에

대한 부분이 주로 많았음을 확인할 수 있다.

2000~2009년에 수행된 융합정보시스템의 핵심 연구의 토픽모델링 결과는 <표 9>와 같다. 논문 9,741개를 통하여 토픽모델링한 결과, 분석/지식/근접, 처리/협업, 예측/물/모바일, 선행/백터/근집, 패턴/인식/시간, 성능/만족/추출, 소셜/인지/시각화, 신경망/적용/감소, 소통/결정/가상, 마이닝/최적/인공으로 결과가 도출되었다. 이는 이 당시의 융합정보시스템 연구로서 정보시스템 분석 기법에 대한 연구가 이뤄졌음을 나타낸다.

〈표 8〉 순수정보시스템 토픽(2000~2009)

Topic	Description	
	Label	Terms
01	정보 기술	model, data, information, design, process, theory, investigation, context, privacy, intelligent, untitled, rules, driven, semantic, technologies, project, requirements, software, research, dss
02	네트워크	business, approach, study, performance, issue, network, quality, special, software, system, consumer, retrieval, factors, case, effects, search, mining, systems, comparison, erp
03	정보	online, time, global, support, strategies, effect, large, outsourcing, industry, erp, real, information, work, architecture, auction, peer, implementation, queries, services, introduction, content
04	인플루언스	analysis, Data, Organizational, adoption, modeling, study, electronic, group, mobile, models, information, influence, developing, approach, mining, efficient, database, strategy, access, security
05	기술	based, web, knowledge, internet, technology, computer, case, role, market, resource, performance, xml, enabled, model, agent, process, empirical, programming, mediated
06	전략	systems, system, development, support, learning, software, perspective, success, effects, strategic, framework, application, management, editorial, services, implementation, approach, collaborative, financial, change
07	공유	information, management, research, trust, support, design, knowledge, networks, supply, sharing, managing, commerce, organizations, chain, distributed, security, framework, exploratory, engineering, processes
08	시장	impact, multi, method, risk, enterprise, evaluation, control, markets, level, multiple, auctions, agent, query, project, dynamic, models, methodology, individual, system, market
09	관계	decision, user, making, service, oriented, relationship, empirical, cross, object, planning, integration, cultural, cognitive, action, process, technical, language, factors, model, end
10	지식	information, technology, commerce, empirical, management, electronic, study, customer, understanding, environment, role, approach, organization, knowledge, social, government, structure, acceptance, case, product

〈표 9〉 융합정보시스템 토픽(2000~2009)

Topic	Description	
	Label	Terms
01	분석, 지식, 근접	analysis, model, knowledge, based, rule, web, case, internet, component, ontology, context, content, text, students, distance, specification, checking, image, nearest, neighbor
02	처리, 협업	systems, information, fuzzy, system, technology, retrieval, expert, theory, database, logic, processing, set, dynamic, rough, reliability, motivation, sets, collaboration, power, routing
03	예측, 룰, 모바일	learning, object, prediction, oriented, modeling, rules, reasoning, service, mobile, language, discovery, commerce, machines, computing, integration, online, acquisition, product, quality, relational
04	선형, 벡터, 군집	computer, programming, user, human, algorithm, clustering, models, regression, interface, estimation, mediated, cost, approach, vector, machine, languages, linear, integer, peer, allocation

〈표 9〉 융합정보시스템 토픽(2000~2009)(계속)

Topic	Description	
	Label	Terms
05	패턴, 인식, 시간	software, design, classification, time, recognition, pattern, development, scheduling, metrics, real, verification, series, program, detection, measurement, image, reliability, stochastic, approximation, maintenance
06	성능, 만족, 추출	performance, selection, feature, problem, multi, agent, memory, group, extraction, validation, method, temporal, spatial, satisfaction, multiple, xml, task, agents, criteria, education
07	소셜, 인지, 시각화	web, process, search, control, social, behavior, quality, query, cognitive, interaction, engineering, patterns, requirements, similarity, index, adaptive, environments, environment, services, visual
08	신경망, 적용, 감소	networks, neural, algorithms, network, genetic, algorithm, models, testing, security, complexity, business, markov, graphs, privacy, personality, study, implementation, architecture, branch, reduction
09	소통, 결정, 가상	management, decision, support, system, communication, making, fault, diagnosis, machine, tree, risk, architecture, systems, project, engineering, learning, detection, virtual, chain, trust
10	마이닝, 최적, 인공	data, mining, optimization, artificial, evaluation, databases, performance, association, distributed, framework, intelligence, intelligent, flow, inference, internet, acceptance, generation, test, differences, wavelet

4.1.4 정보시스템 토픽(2010~2015)

2010~2015년에 연구된 순수정보시스템 논문 3,535개에 대한 LDA 결과는 제품 서비스, 커뮤니티, 융합, 온톨로지, 데이터, 정보, 경쟁, 시장, 최적화 그리고 모바일 기술이라는 10가지 주제로 도출되었다. <표 10>에서의 토픽모델링 결과는 최근 비즈니스와 연관된 다양한 학문들이 정보시스템 학문과 융합을 이루어 새로운 연구와 기술을 창출해내고 있다는 점을 보여준다. 그리고 경쟁 시장에 있어 제품과 서비스에 대한 발전뿐만 아니라 최적화에 중점을 둔 연구가 진행되고 있

다는 것도 알 수 있다.

<표 10>의 순수정보시스템 토픽모델링 결과와 동시대에 이뤄진 14,506개의 융합시스템 연구를 기반으로 한 결과는 결정/시간/관리, 모바일/보안/추론, 세분화/감지, 신경망/실증, 기계학습, 예측/위험/상호작용, 인식/질/다이나믹, 프라이버시/채택, 분류/인공/검증, 진화/메모리/군집으로 10가지 토픽들이 <표 11>과 같이 산출되었다. 이는 정보시스템과 공학적인 학문들이 융합하여 정보시스템 학문에 있어, 기술적인 연구 동향이 진행되었음을 알 수 있다.

〈표 10〉 순수정보시스템 토픽(2010~2015)

Topic	Description	
	Label	Terms
01	제품, 서비스	service, software, process, design, quality, impact, context, product, oriented, strategic, special, networks, innovation, electronic, services, aware, media, industry, network, issues

<표 10> 순수정보시스템 토픽(2010~2015)(계속)

Topic	Description	
	Label	Terms
02	커뮤니티	support, knowledge, investigation, risk, sharing, search, systems, governance, study, framework, communities, success, introduction, large, exploring, community, network, methodology, markets, detection
03	융합	multi, influence, multiple, issue, project, research, perspective, level, criteria, outsourcing, group, privacy, relational, theoretical, structure, dimensional, firm, hybrid, high, projects
04	온톨로지	based, decision, case, making, web, knowledge, content, role, time, cost, network, internet, ontology, resource, event, real, distributed, agent, effective, usage
05	데이터	data, model, research, system, online, effects, performance, theory, models, management, development, method, learning, social, factors, special, relationship, conceptual, firms, implementation
06	정보	information, empirical, adoption, online, approach, strategy, efficient, perspective, services, system, electronic, learning, practice, acceptance, word, introduction, managing, framework, evaluation, implementation
07	경쟁	systems, approach, supply, open, performance, digital, source, mining, management, commerce, social, reviews, chain, quality, capabilities, perspective, issue, competition, processes, determinants
08	시장	analysis, business, social, organizational, consumer, cross, pricing, firm, information, framework, market, model, strategies, security, investment, competitive, virtual, approach, communication, cultural
09	최적화	information, systems, study, modeling, management, perspective, behavior, search, network, role, government, web, understanding, security, impact, optimal, social, privacy, requirements, implications
10	모바일 기술	technology, user, trust, impact, effect, public, enabled, media, rfid, computer, review, untitled, enterprise, virtual, study, work, mediated, mobile, continuance, mobile

<표 11> 융합정보시스템 토픽(2010~2015)

Topic	Description	
	Label	Terms
01	결정, 시간, 관리	system, decision, support, time, making, social, web, management, series, expert, context, identification, collaborative, real, allocation, domain, design, estimation, sequence, awareness
02	모바일, 보안, 추론	optimization, design, selection, models, feature, acceptance, patterns, extraction, mobile, security, reduction, representation, simulation, inference, commerce, integration, discovery, uncertainty, recommender, face
03	세분화, 감지	model, knowledge, based, web, detection, virtual, service, ontology, rule, search, diagnosis, fault, segmentation, distance, game, market, data, oriented, case, graph
04	신경망, 실증	analysis, networks, network, data, neural, behavior, image, mining, generation, random, test, language, empirical, consumer, evaluation, factor, idea, discriminant, word, communication

〈표 11〉 융합정보시스템 토픽(2010~2015)(계속)

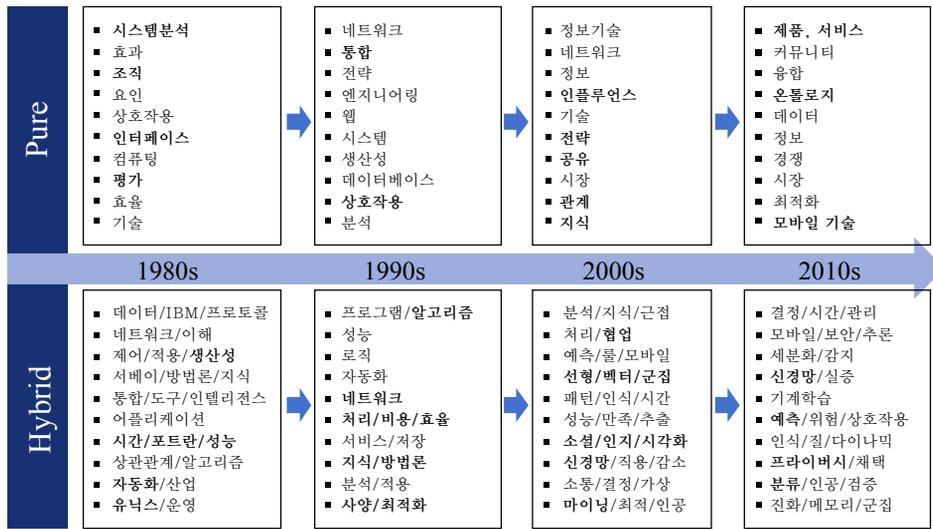
Topic	Description	
	Label	Terms
05	기계학습	fuzzy, computer, machine, programming, sets, vector, swarm, data, human, logic, communication, trust, particle, machines, mediated, motivation, rough, cost, ahp, usage
06	예측, 위험, 상호작용	software, prediction, online, mining, internet, social, risk, development, management, text, interaction, evaluation, scheduling, transform, games, customer, assessment, business, aggregation
07	인식, 질, 다이나믹	systems, information, technology, management, process, quality, control, framework, cognitive, retrieval, dynamic, personality, health, methods, life, processing, power, load, database, task
08	프라이버시, 채택	performance, multi, method, multiple, regression, product, satisfaction, privacy, criteria, group, vector, agent, adoption, ranking, perspective, attribute, objective, sensor, agents, expression
09	분류, 인공, 검증	classification, theory, user, recognition, set, artificial, pattern, students, services, chain, gender, impact, perceived, testing, differences, architecture, education, efficiency, supply, validation
10	진화, 메모리, 군집	algorithm, learning, algorithms, genetic, search, problem, clustering, rules, approach, media, environments, databases, query, evolution, computing, similarity, function, modeling, memory, wavelet

4.1.5 정보시스템 토픽(1980~2015)

<그림 3>을 보면 1980년대는 모든 기업이 당면하는 의사결정문제에 있어, 더욱 용이하고 효과적인 해결을 돕는 “의사결정 지원시스템(Decision Support System)” 등의 다양한 컴퓨팅 기술이 발전했다. 이처럼 당시에는 시스템 분석 체계에 초점을 둔 연구가 활발하게 진행되었다. 1990년대는 “인터넷의 상용화” 됨으로써, “웹(Web)”이라는 가상 환경이 구축되었다. 동시대에 기업의 비즈니스가 점점 오프라인에서 온라인으로 변화하면서 다양한 의견 수렴 및 비교가 가능하게 되었다. 이를 바탕으로 기업의 이윤을 최대화 창출하기 위한 IT(Information Technology) 전략수립과 IT 기반 조직의 생산성 증대를 위한 연구가 진행되었다.

2000년대는 더욱 고급화된 정보화 기술을 통해서 IT 기반의 지식 관리가 용이하며, 시장의 유용

한 정보와 지식 그리고 전략을 도출할 수 있는 “데이터 마이닝(Data Mining)”에 대한 연구가 활성화되었다. 마지막으로 2010년대를 보면, 스마트 비즈니스 시대 도래로 각종 모바일 기술과 소셜 미디어가 등장하였다. 이를 통해, 각종 모바일 로그 데이터, 소셜 미디어 데이터 등 실시간으로 데이터가 폭증하는 빅데이터 세상이 출현하였다. 이로 인해 엄청난 양의 데이터 저장소를 구축할 수 있는 인터넷 상의 서버, “클라우드 컴퓨팅(Cloud Computing)” 기술이 발달했고, 이것이 곧 스마트 비즈니스 시대의 빅데이터 기술을 견인해 준 것으로 확인할 수 있다. 나아가, 최근에 들어서는 정보화 시대의 시장 경쟁력이 무조건적인 기술 발전이 아닌, 고객 만족을 최우선으로 하면서 고객 욕구를 충족시킬 수 있는 최적화된 제품과 서비스가 무엇인가가 이슈가 되며 연구가 진행되고 있음을 추론할 수 있다.



<그림 3> 시기 변화에 따른 순수 정보시스템과 융합정보시스템 토픽의 변화(1980~2015)

## 4.2 네트워크 분석을 기반으로 한 정보시스템 학문의 동향 분석

본 장에서는 기존에 연구되어온 정보시스템 학문의 논문 데이터를 통해, 논문들이 가지고 있는 여러 가지 속성(예: 제목, 키워드, 저자명, 기간 등)을 활용하여 네트워크를 가시화하고, 특징적 관계를 분석하고자 한다. 먼저, 1990년대부터 2015년까지의 기간을 ‘1990~1999’, ‘2000~2009’, ‘2010~2015’으로 세분화하였고, 각 연구의 키워드를 중심으로 한 네트워크 분석을 통해 정보시스템 학문의 동향을 파악하였다.

### 4.2.1 네트워크 분석(1990~1999)

이 절에서는 1990~1999년대, 약 10년 간 정보시스템 분야에서 연구되어온 논문을 중심으로 한 네트워크 분석 결과를 이야기하고자 한다. 먼저, Chan et al.(2014)의 연구를 활용해 정보시스템 논문의 갈래를 두 가지 분야(Pure IS와 Hybrid IS)로 분류한 후 논문의 키워드 중심, 논문의 키워드와 저자명 중심으로 네트워크 분석을 시행하였다. 시각화된 결과는 <그림 4> 및 <그림 5>와 같다.

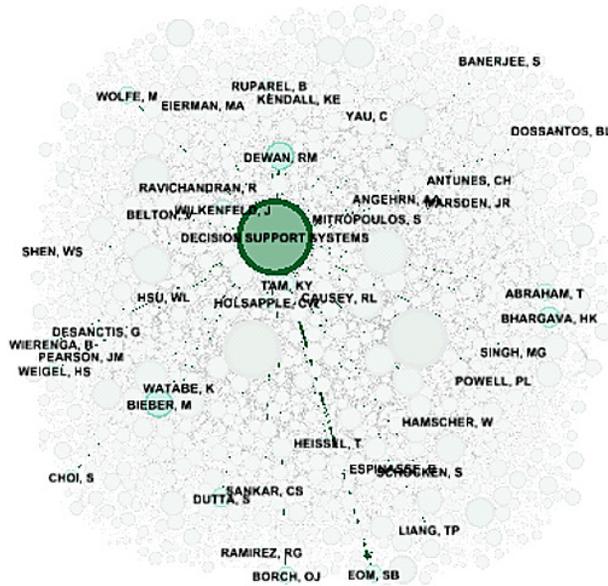
분석 결과, 첫째로 순수정보시스템 분야에서는 동 기간 내 3,575개의 키워드가 도출되었고, 키워드와 저자명 간에는 5,556개의 연결고리를 발견할 수 있었다. 둘째, 융합정보시스템 분야에서는 동 기간 내 6,265개의 키워드가 도출되었으며, 키워드와 저자명 간에는 10,730개의 연결고리를 발견할 수 있었다.

위 결과를 비교해 보았을 때, 비교적 정통적인 정보시스템 학문을 연구하는 순수정보시스템 분야에 비해, 타 학문과 정보시스템 학문을 융복합하여 연구하는 융합정보시스템 분야가 훨씬 많은 화두와 키워드를 가지고 활발하게 연구되고 있음을 알 수 있었다. 또한, 순수정보시스템 분야에서는 키워드와 저자명 간의 연결고리가 5,556개 발견된 것에 비해, 융합정보시스템 분야에서는 키워드와 저자명 간의 연결 고리가 10,730개 연결되어 있는 것으로 보아, 연구자들이 융합정보시스템 분야에서 훨씬 활발한 아이디어를 가지고 연구하고 있는 것으로 분석되었다. 그러므로 순수정보시스템 분야보다 타 학문과 융복합하여 다양한 소재의 학문들이 연구되고 있는 융합정보시스템이 연구자들에게 연구 대상으로서 많은 각광을

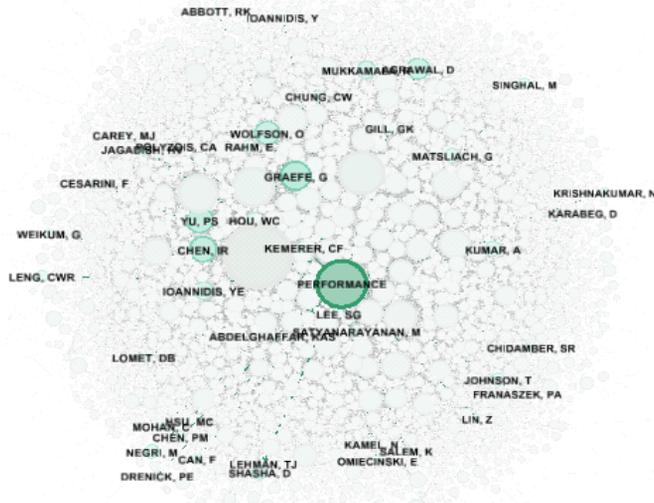
받고 있다고 말할 수 있다.

<그림 4>를 보면, “Decision Support Systems(의사결정 지원 시스템)”이라는 키워드를 중심으로 각 연구자들이 분포되어 있는 것을 확인할 수 있으며, <그림 5>에서는 “Performance(실적, 성과)”

관련 키워드를 중심으로 많은 연구자들이 네트워크를 형성하여 이들을 중심으로 관련 연구들이 행해지고 있음을 알 수 있다. 실제로 1990년대부터 기업에서는 단순한 데이터 처리, 거래 처리 시스템을 떠나 DSS(Decision Support System)라는

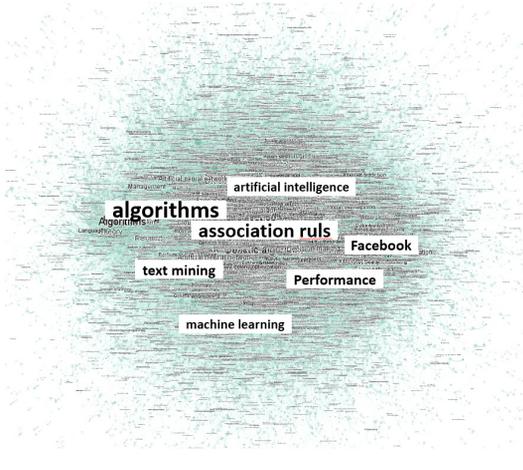


<그림 4> 순수정보시스템 1990~1999 키워드



<그림 5> 융합정보시스템 1990~1999 키워드

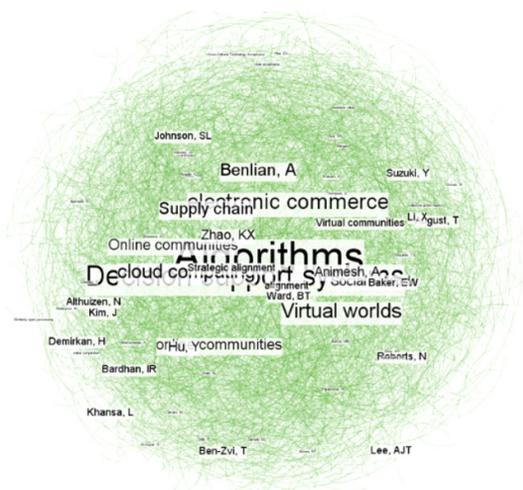




<그림 7> 융합정보시스템 2000~2009 키워드

#### 4.2.3 네트워크 분석(2010~2015)

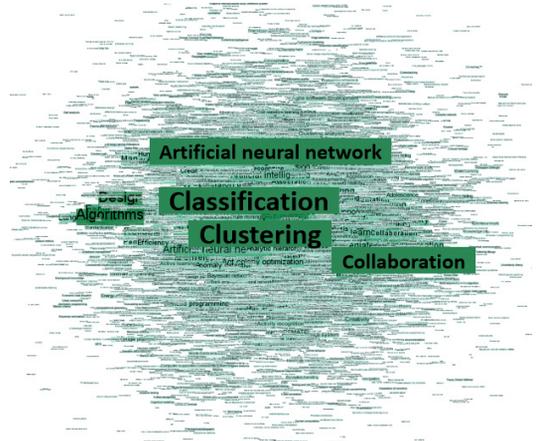
본 절에서는 2010년부터 2015년까지 순수정보시스템 분야에서 다루어진 논문의 키워드를 네트워크 분석함으로써 특징적으로 도출된 결과를 논의해 보고자 한다. <그림 8>과 같이, 2010년에 들어 오면서부터 순수정보시스템 학문에서 주로 다루어진 키워드로는 “Algorithms(알고리즘)”, “Decision support(의사결정 지원)”, “Cloud Computing(클라우드 컴퓨팅)”, “E-commerce(전자 상거래)”, “Virtual



<그림 8> 순수정보시스템 2010~2015 키워드

World(가상 세계)”, “Virtual Communities(가상 공동체)”, “Online Communities(온라인 공동체)”가 대표적이었다. 특징적으로, “Decision support(의사결정 지원)”, “E-commerce(전자 상거래)”는 2000년대부터 활발히 연구되어왔고, 지금까지도 많은 연구가 되어오는 이슈이기에 큰 관심을 받았다는 점을 확인할 수 있다.

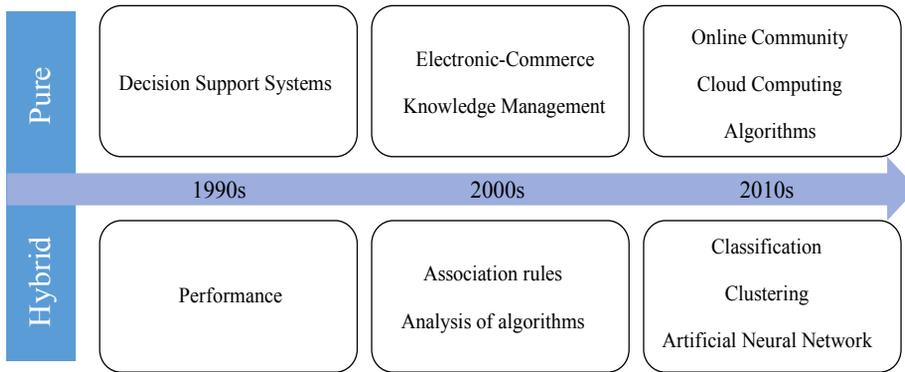
또한, 융합정보시스템 네트워크 분석을 시행하였을 때, “Classification(분류화)”, “Clustering(군집화)”, “Algorithms(알고리즘)”, “Design(디자인)”, “Artificial neural network(인공 신경망)”, “Optimization(최적화)”, “Collaboration(협업)”과 같은 키워드가 중심적으로 나타났음을 <그림 9>에서 볼 수 있다. 2000년대 초반의 융합정보시스템 분석 시 등장한 키워드 구성과 크게 다른 점은, 인공신경망과 협업과 같은 키워드의 새로운 등장이다. 다른 학문 분야와의 협업에 대한 관심이 고조되었으며, 인공신경망에 대한 연구 수요도 함께 증가하였음을 확인할 수 있다.



<그림 9> 융합정보시스템 2010~2015 키워드

#### 4.2.4 네트워크 분석(1990~2015)

1990년도 정보시스템 학문은 의사결정 지원 시스템에 대한 연구를 진행하였으며, 그로 인해 다양한 학문들이 접목되어 “어떻게 의사결정 지



<그림 10> 시기 변화에 따른 순수정보시스템과 융합정보시스템의 핵심 연구 비교(1990~2015)

원 시스템을 연결하여, 더욱 큰 효율성과 이윤을 얻을 수 있는가?”에 대해 의문을 가지고 활발한 연구를 진행했다. 2000년대에 들어와서는 데이터 마이닝(Data Mining)이 정보시스템 학문에서 큰 화두가 되면서, 다양한 학문을 데이터 마이닝을 통해 분석, 조금 더 현명하고 기발한 통찰력을 도출하기 위한 연구가 증가하였다. 2010년 이후에는 점점 우리의 생활이 스마트화 되어 데이터 양이 급증하고, 빅데이터 분석 기술 또한 더욱 발달함에 따라 “클라우드 컴퓨팅 기술”도 함께 발달하였다. 이로 인하여 비정형 정보, 정형 정보 등 다양화된 정보를 더욱 일목요연하게 분류하고 군집화하는 classification(분류화)과 clustering(군집화) 기법이 급속도로 발전하기 시작했다. 이상의 내용은 <그림 10>과 같다.

#### 4.3 네트워크 분석을 기반으로 한 공동 연구 네트워크 분석

본 장에서는 정보시스템의 연구 동향뿐만 아니라, 정보시스템 연구의 중심에 서서 정보시스템 연구의 흐름을 이끄는 허브 역할의 연구자를 찾고자 하였다. 따라서 위에서 분류한 순수정보시스템 카테고리의 저널에서 2010년부터 2015년까지 2편 이상 논문을 투고한 총 2,327명의 연구자를 대상으로 저자명과 논문 데이터를 활용하여

공동 연구 네트워크 분석을 진행하였다. 그 결과, 2,327개의 노드(저자명)와, 5,831개의 연결관계(논문)가 형성되었으며, 그 중심에서 나타난 연구자는 “Zhang, J.”, “Benbasat, I.”, “Lowly, P. B.”이다. 결과로 나타난 3명의 연구자들은 다양한 정보시스템 연구에서의 중추적 역할을 맡아 정보시스템 연구를 이끌고 있음을 알 수 있다. 실제로, “Lowly, P. B.”의 경우, 약 70개의 정보시스템 연구 관련 저널에 150개의 정보시스템 논문을 투고하여 다양한 학문분야에서 정보시스템의 연구가 성행할 수 있도록 기여 하였다.

### V. 결론 및 시사점

본 연구는 기존의 정보시스템 분야에서 연구자들이 어떠한 주제로 연구해왔는지에 대한 흐름을 파악하고자 하였다. 특히, 융합과학이 각광받기 이전의 전통적인 정보시스템 분야, 즉, 순수정보시스템 분야의 저널과 융합정보시스템 분야의 저널에 대하여 토픽모델링 분석을 시행함으로써 두 카테고리의 정보시스템 분야의 주요 키워드 분석을 할 수 있는 기반을 마련하고자 하였다. 그 결과, 정보시스템 연구주체는 1980년도에는 효율성과 시스템에 관련된 주제들에 관한 연구들이 많았지만 점차 알고리즘, 제품/서비스 방향으로 변화하였음을 알 수 있다. 네트워크에 관련된 주

제는 1980년대부터 꾸준히 연구되었으며, 알고리즘과 같은 분석 기법과 관련된 주제는 시간이 지남에 따라 각광을 받고 있다는 점도 연구 결과를 통해 확인할 수 있었다.

또한, 본 연구는 네트워크 분석을 통해 순수정보시스템 및 융합정보시스템 분야에서의 관심분야가 어떠한 방향성을 갖고 이동하는지에 대한 분석도 실시하였다. 그 결과, 융합정보시스템 분야에서 1990년대에는 의사결정 시스템에 대한 연구가 중심을 이루었다면, 2000년대로 넘어오면서부터 보다 융합적 측면을 강조하는 텍스트 마이닝, 기계학습과 같은 주제가 각광받기 시작했음을 알 수 있었다. 나아가, 2010년 이후에는 군집화, 협업 등과 같은 주요 키워드가 새롭게 등장하면서 융합적 측면과 분화적 측면이 동시에 주요 연구 주제로 다루어져 왔음을 확인하였다. 이는 비즈니스와 연관된 순수정보시스템 연구에 공학과 같은 기술적인 학문이 단계적으로 추가 및 응용되는 변화를 말해준다. 즉, 순수정보시스템과 융합정보시스템의 주제들이 비슷한 맥락과 방향성을 가지고 있지만 순수정보시스템과 융합정보시스템의 주제들의 범위와 깊이에는 차이가 있음을 알 수 있다. 마지막으로 본 연구의 분석 결과를 통해 정보시스템 분야에서의 학문적 연구의 중심에 서서 계획하는 연구자들이 누구이며, 그들을 중심으로 생성되는 정보시스템 연구 동향 및 흐름을 파악함으로써 앞으로의 연구 방향성을 확인할 수 있기를 기대한다.

정보시스템 학문의 동향을 살펴보고자 한 본 연구에서 도출된 결과를 기반으로 다음과 같은 시사점을 제시할 수 있다. 첫째, 이전까지는 IS 분야의 연구 동향에 대해 토픽모델링, 네트워크 분석을 통한 다양한 연구가 부족했다. 이에 본 연구에서는 기존의 정보시스템 논문에서 다루어졌던 주요 주제들을 파악하는 것에서 더 나아가, 네트워크 분석을 통한 데이터 시각화로 향후 정보 동향을 제시하였다. 둘째, Chan *et al.*(2015)의 연구를 활용해 정보시스템 분야를 순수정보시스템

(Pure IS)과 정보시스템과 타 학문이 융합된 융합정보시스템(Hybrid IS)이라는 두 개의 분야로 나누어 분석하는 새로운 분석방법을 채택하였다는 점에서 의의가 있다. 과거에 정보시스템학에서 주로 다루었던 주제들을 순수정보시스템으로 구분하고, 융합과학이 대두되고 있는 최근 동향에 맞추어 타 학문과 융합된 융합정보시스템 분야를 별도로 분석하였다는 점에서 그 의미를 찾을 수 있다. 셋째, 이전 정보시스템 분야의 연구 발전 과정 및 향후 해당 분야의 발전 전망성을 예측해 볼 수 있는 연구라 할 수 있다. 특히, 본 연구의 결과 정보시스템 분야는 융합과학(Convergence Science)의 특징을 더 갖고 있음을 알 수 있다. 이는 앞으로 정보시스템 분야에서 새로운 연구를 진행할 다방면의 연구자들에게 연구 주제를 선정함에 있어 보다 명확한 방향성을 제시해 줄 수 있을 것으로 기대한다. 넷째, 본 연구는 정보시스템 분야에서 다루는 주제들이 실무와 얼마나 밀접한 관계를 가져왔는지에 대한 분석을 통해, 정보시스템 분야에서 현재 관심을 가지는 주제들에 대한 통찰력을 제공할 수 있다. 마지막으로, 정보시스템 연구의 허브역할을 하는 연구자들을 분석함으로써, 공동 저자 네트워크를 통한 정보시스템 연구의 확장에 대한 시사점을 제공하였다.

향후 본 연구와 관련된 연구를 진행함에 앞서 다음과 같은 향후 연구들을 제안하고자 한다. 첫째, 저널들 간의 Impact Factor를 적용한 기존 선행 연구인 Chan *et al.*(2006)의 연구를 적용하여 가중치를 적용한 연구를 진행하고자 한다. 둘째, 인용 수를 고려하여 연구 주제와 연구자들의 영향력에 대한 분석도 진행하고자 한다. 셋째, CIO(Chief of Information Officer) 저널들의 주제와 키워드들 간의 비교를 통해, 정보시스템의 연구와 실무적 관계에 있어 정보시스템이 어떠한 역할을 해왔는지에 대한 분석을 추가하고자 한다. 마지막으로, 과거에 정보시스템 연구의 주제를 시간에 따라 더욱 세분화된 분석을 진행하여 과거에 소멸된 주제가 다른 주제와 결합되어 나올 수 있음을 연구하고자 한다.

## 참고 문헌

- [1] 강범일, 송 민, 조화순, “토픽모델링을 이용한 신문 자료의 오피니언 마이닝에 대한 연구”, *한국문헌정보학회지*, 제47권, 제4호, 2013, pp. 315-334.
- [2] 강성경, 유 환, 이영재, “텍스트 마이닝과 소셜 네트워크 분석을 이용한 재난대응 용어분석”, *Information Systems Review*, 제18권, 제1호, 2016, pp. 141-155.
- [3] 박자현, 송 민, “토픽모델링을 활용한 국내 문헌 정보학 연구 동향 분석”, *정보관리학회지*, 제30권, 제1호, 2013, pp. 7-32.
- [4] 안정국, 김희웅, “집단지성을 이용한 한글 감성어 사전 구축”, *지능정보연구*, 제21권, 제2호, 2015, pp. 49-67.
- [5] 이 석, 정병규, “IEEE ICRA 2011~2014 년 논문을 활용한 로봇 및 제어자동화 분야 연구동향 분석”, *한국정밀공학학회지*, 제32권, 제3호, 2015, pp. 233-241.
- [6] 정근하, “텍스트마이닝과 네트워크 분석을 활용한 미래예측방법 연구”, *한국과학기술평가원 연구보고서*, 2011.
- [7] 조승연, 최지은, 이규현, 김희웅, “고객 온라인 구매후기를 활용한 추천시스템 개발 및 적용”, *Information Systems Review*, 제17권, 제3호, 2015, pp. 77-93.
- [8] 최기선, 함영균, 임경태, 박정열, 윤용운, “Linked Data를 위한 한국어 자연언어처리 플랫폼”, *제24회 한글 및 한국어 정보처리 학술대회*, 2012, pp. 16-20.
- [9] Blei, D. M., “Probabilistic topic models”, *Communications of the ACM*, Vol.55, No.4, 2012, pp. 77-84.
- [10] Blei, D. M., A. Y. Ng, and, M. I. Jordan, “Latent dirichlet allocation”, *Journal of Machine Learning Research*, Vol.3, 2003, pp. 993-1022.
- [11] Blei, D. M., T. L. Griffiths, and M. I. Jordan, “The nested Chinese restaurant process and Byesian nonparametric inference of topic hierarchies”, *Journal of the ACM(JACM)*, Vol.57, No.2, 2010, pp. 111-142.
- [12] Bonacich, P., “Power and Centrality: A Family of Measures”, *American Journal of Sociology*, Vol.92, No.5, 1987, pp. 1170-1182.
- [13] Chan, C. H., H. W. Kim, and W. C. Tan, “Information systems citation patterns from international conference on information systems articles”, *Journal of the Association for Information Science and Technology*, Vol.57, No.9, 2006, pp. 1263-1274.
- [14] Chan, H. C., V. Guness, and H. W. Kim, “A method for identifying journals in a discipline: An application to information systems”, *Information & Management*, Vol.52, No.2, 2015, pp. 239-246.
- [15] Chen, W. and R. Hirschheim”, A paradigmatic and methodological examination of information systems research from 1991 to 2001”, *Information Systems Journal*, Vol.14, No.3, 2004, pp. 197-235.
- [16] Fischer, H. R., “Abductive reasoning as a way of worldmaking”, *Foundations of Science*, Vol.6, No.4, 2001, pp. 361-383.
- [17] Galaskiewicz, J. and S. Wasserman, “Social network analysis concepts, methodology, and directions for the 1990s”, *Sociological Methods & Research*, Vol.22, No.1, 1993, pp. 3-22.
- [18] Gerrish, S. and D. M. Blei, “A language-based approach to measuring scholarly impact”, *The 27th International Conference on Machine Learning*, Vol.10, 2010, pp. 375-382.
- [19] Grimmer, J., “A bayesian hierarchical topic model for political texts: Measuring expressed agendas in senate press releases”, *Political Analysis*, Vol.18, No.1, 2010, pp. 1-35.
- [20] Kam, M. and M. Song, “A study on differences

- of contents and tones of arguments among newspapers using text mining analysis”, *Journal of Intelligence and Information Systems*, Vol.18, No.3, 2012, pp. 53-77.
- [21] Losee, J., *A Historical Introduction to the Philosophy of Science*, Oxford University Press, London, 1972.
- [22] Mimno, D., H. M. Wallach, and A. McCallum, “Gibbs Sampling for Logistic Normal Topic Models with Graph-Based Priors”, 2008, Available at <https://people.cs.umass.edu/~wallach/publications/mimno08gibbs.pdf>.
- [23] Négyessy, L., T. Nepusz, L. Zalányi, and F. Bazsó, “Convergence and divergence are mostly reciprocated properties of the connections in the network of cortical areas”, *Proceedings of the Royal Society of London B: Biological Sciences*, Vol.275, No.1649, 2008, pp. 2403-2410.
- [24] Newman, D. J. and S. Block, “Probabilistic topic decomposition of an eighteenth-century american newspaper”, *Journal of the American Society for Information Science and Technology*, Vol.57, No.6, 2006, pp. 753-767.
- [25] Song, M. and S. Y. Kim, “Detecting the knowledge structure of bio informatics by mining full-text collections”, *Scientometrics*, Vol.96, No.1, 2013, pp. 183-201.
- [26] Song, M., S. Kim, G. Zhang, Y. Ding, and T. Chambers, “Productivity and influence in bioinformatics: A bibliometric analysis using PubMed central”, *Journal of the Association for Information Science and Technology*, Vol.65, No.2, 2014, pp. 352-371.
- [27] Steyvers, M. and T. Griffiths, “Probabilistic topic models”, *Handbook of Latent Semantic Analysis*, Vol.427, No.7, 2007, pp. 424-440.
- [28] Walstrom, K. A. and B. C., Hardgrave”, Forums for information systems scholars: III”, *Information & Management*, Vol.39, No.2, 2006, pp. 117-124.
- [29] Zhao, W. X., J. Jiang, J. Weng, J. He, E. P. Lim, H. Yan, and X. Li, “Comparing twitter and traditional media using topic models”, *Advances in Information Retrieval*, April 2011, pp. 338-349.

Information Systems Review

Volume 18 Number 3

September 2016

## Exploring Dynamics of Information Systems Research Trend Using Text Mining Approach

Jungkook An\* · Sodam Kim\*\* · Hee-Woong Kim\*\*\*

### Abstract

Recent research on information and communication technology and Internet-of-Things indicates that convergence and integration facilitate the development of various technologies. Similarly, related academic theories and technologies have also gained attention. This paradigm shift facilitated the convergence and integration of academic disciplines. In particular, information systems have become initiators of change. However, only a limited number of studies have been conducted on information systems. To address this gap, this study explores the future direction of information systems based on the core concepts and results of the comparative analysis conducted on research trends. We considered 48,102 data obtained from international top journals from 1980 to 2015. We analyzed journal titles, authors, abstracts, and keywords. We conducted the network analysis on existing collaborative studies and performed comparative analysis to visualize the results. The results provide an in-depth understanding of information systems and provides directions for future research on this area.

**Keywords:** *Information Systems, Research Trends, Collaborative Research Network, Topic Modeling, Network Analysis*

---

\* Graduate School of Information, Yonsei University

\*\* Graduate School of Information, Yonsei University

\*\*\* Corresponding Author, Professor, Graduate School of Information, Yonsei University

## ◎ 저 자 소 개 ◎



**안 정 국 (jace@yonsei.ac.kr)**

현재 연세대학교 정보대학원에서 박사과정 재학 중이며, 주요 연구분야는 Big Data Analytics, Natural Language Processing, Business Strategy, Social media marketing, Brand engagement 등이다.



**김 소 담 (ksodam@yonsei.ac.kr)**

현재 연세대학교 정보대학원에서 박사과정 재학 중이며, 주요 연구분야는 Digital Business, Social Network Services, Online Communities, Social Media/Online Marketing 등이다.



**김 희 웅 (kimhw@yonsei.ac.kr)**

National University of Singapore 정보시스템학과에서 근무 후, 현재 연세대학교 정보대학원 교수로 근무 중이다. 주요 관심분야는 디지털 비즈니스, 정보시스템 관리 및 활용 등이다. 관련 연구들은 MIS Quarterly, Information Systems Research, Journal of Management Information Systems, Journal of the Association for Information Systems, Journal of Retailing 등의 저널에 40여 편의 논문이 게재되었다. Journal of the AIS, IEEE Transactions on Engineering Management 저널의 편집위원을 역임하였다.

논문접수일 : 2016년 05월 04일

게재확정일 : 2016년 08월 22일

1차 수정일 : 2016년 06월 17일