

국내 산업공학 연구 주제 2001~2015

정보권¹ · 이학연^{2*}

¹서울과학기술대학교 IT정책전문대학원 / ²서울과학기술대학교 글로벌융합산업공학과

Research Topics in Industrial Engineering 2001~2015

Bokwon Jeong¹ · Hakyon Lee²

¹The Graduate School of Public Policy and Information Technology,
Seoul National University of Science and Technology

²Department of Industrial and Systems Engineering, Seoul National University of Science and Technology

Over the last four decades, industrial engineering (IE) research in Korea has continued to evolve and expand to respond to social needs. This paper aims to identify research topics in IE research and explore their dynamic changes over time. The topic modeling approach, which automatically discovers topics that pervade a large and unstructured collection of documents, is adopted to identify research topics in domestic IE research. 1,242 articles published from 2001 to 2015 in two IE journals issued by the Korean Institute of Industrial Engineers were collected and their English abstracts were analyzed. Applying the Latent Dirichlet Allocation model led us to uncover 50 topics of domestic IE research. The top 10 most popular topics are revealed, and topic trends are explored by examining the dynamic changes over time. The four topics, technology management, financial engineering, data mining (supervised learning), efficiency analysis, are selected as hot topics while several traditional topics related with manufacturing are revealed as cold topics. The findings are expected to provide fruitful implications for IE researchers.

Keywords: Industrial Engineering, Research Topic, Text-mining, Topic model, Latent Dirichlet Allocation (LDA)

1. 서론

대부분의 공학 분야는 산업 및 기술의 변화 추세에 따라 적응하고 시대적 요구에 발맞춰 발전한다. 산업공학은 프레더릭 테일러(Frederick Taylor)의 과학적 관리(scientific management)를 시작으로, 헨리 간트(Henry Gantt)와 길브레스(Gilbreth) 부부 등의 연구를 통해 그 초기 형태가 만들어졌으며(Salvendy, 2001), 이후 생산 및 품질 관리에 대한 연구가 본격적으로 이루어지기 시작했다. 제2차 세계대전을 전후한 OR(operations research) 분야의 발전으로 본격적인 케도에 오른 산업공학 분야는 1980년대 이후에는 컴퓨터의 발전 및 확산에 따라 정보시스템 분야까지 그 범위를 확장하였으며, 최근 들어서는 금융

공학, 서비스공학, 기술경영 등을 아우르며 오늘날의 산업시스템공학, 산업경영공학에 이르렀다(Park, 2015). 국내에서는 1974년 대한산업공학회가 설립되어 지난 40여 년간 학술지 발간 및 다양한 연구 활동과 사업 등을 통해 국내 산업공학 발전에 핵심적인 역할을 수행하고 있으며, 현재 70여 개의 대학에 설치된 산업공학 관련 학과를 중심으로 활발한 산학연 연구를 통해 외연을 확장하고 있다.

이처럼 사회적 발전과 변화에 따라 산업공학 분야의 발전이 함께 이루어지고 있다. 가령 최근의 사물인터넷, 빅데이터, 인공지능 등의 기술들의 보급과 적용이 세계적인 이슈가 되면서 이들 기술들이 적용되는 산업 전반의 모든 영역에서 산업공학의 역할과 필요성이 함께 높아질 것이다. 따라서 산업공학이

이 연구는 서울과학기술대학교 교내연구비 지원으로 수행되었습니다(2016-0646).

* 연락저자 : 이학연 교수, 01811 서울시 노원구 공릉로 232 서울과학기술대학교, Tel : 02-970-6469, Fax : 02-974-2849, E-mail : hylee@seoultech.ac.kr
2016년 9월 6일 접수; 2016년 10월 10일 수정본 접수; 2016년 11월 28일 게재 확정.

새로운 기술과 영역을 수용하고 산업 현장에 기여하기 위해서는 현시점의 산업공학 현황을 면밀하게 파악하여 산업공학 본연의 목적대로 산업 현장에 효율적이고 가치 있는 기여를 할 수 있는 방안을 미리 준비하고 개발해야 할 것이다.

한편 산업공학 분야의 지속적인 발전이 이루어짐에 따라, 그간 산업공학 연구의 발전 과정을 분석하고 현재 연구 동향을 파악함으로써 향후 산업공학 연구의 비전 및 발전 방향을 제시하려는 시도가 이루어져 왔다. 대한산업공학회지에서는 대한산업공학회 40주년 기념 특집호를 통해, 경영과학(Kim *et al.*, 2014), 품질 및 신뢰성(Yum *et al.*, 2014), 인간공학(Chung *et al.*, 2014), 제조시스템공학(Choi *et al.*, 2014), 생산관리(Park *et al.*, 2014) 등 산업공학의 세부 분야별로 역사 및 전망에 대한 긍정적인 고찰이 이루어졌다. 실증 데이터를 이용한 정량적인 분석으로는 산업공학 분야의 대표적인 국제학술지인 IIE Transactions을 대상으로 학술지 논문 초록에 포함된 키워드를 분석하여 산업공학 연구의 현황과 추이를 관찰한 연구(Cho and Kim, 2012), 대한산업공학회지, IE Interface, 한국산업경영시스템학회지, 한국경영과학회지 등 4개의 국내 학술지에 게재된 논문들에 활용된 연구 기법을 분석한 연구(Cho *et al.*, 2014), WoS(Web of Science)의 산업공학 분야에 속한 43개 산업공학 국제 학술지 간의 인용 관계를 바탕으로 산업공학과 타 학문과의 융합 관계를 밝히는 연구(Jeong and Lee, 2016) 등이 수행되었다.

이들 연구는 산업공학 분야의 논문 데이터를 분석하여 시사점을 도출함으로써 산업공학 연구 현황을 실증적으로 파악했다는 의의가 있지만, 향후 산업공학 분야의 유망 연구 주제 설정 및 발전 방향 수립을 위한 보다 구체적인 시사점을 제공하기 위해서는 산업공학 분야 내의 세부 연구 주제들의 내용 및 분포를 파악하고 이들의 동적 변화를 분석할 필요가 있다. 특정 도메인 내의 세부 연구 주제들에 대한 탐색은 대부분의 학문 분야에서 이루어져 왔지만(Ramos-Rodriguez and Ruiz-Navarro, 2004), 산업공학 분야에 대한 이러한 연구는 찾아보기 어렵다. 이에 본 연구에서는 국내 산업공학 논문을 분석하여 산업공학의 세부 주제들이 어떻게 구성되어있는지, 핵심 연구 주제는 무엇인지, 최근 들어 상승 및 하락하고 있는 연구 주제는 무엇인지를 파악한다.

논문 정보를 활용하여 학술 분야별 연구 주제를 파악하기 위한 목적으로, 내용 분석(contents analysis), 계량서지 분석(bibliometric analysis), 텍스트 마이닝(text-mining) 등의 정량적인 분석 방법이 활용되고 있다. 보다 구체적으로는 동시출현 단어분석(co-word analysis)과 동시인용분석(co-citation analysis)이 학술 도메인 내의 연구 영역(research front) 탐색에 활용되어 온 대표적인 기법이다(Cobo *et al.*, 2011). 동시출현단어분석은 키워드 수준의 접근 방법으로, 텍스트 마이닝을 바탕으로 논문 내 키워드를 추출하여 여러 논문에서 두 키워드가 동시에 출현한 빈도수를 바탕으로 네트워크를 구축하고, 키워드들을 군집화 하여 주제를 정의한다(Callon, 1983). 동시인용분석은 두 논문이 다른 논문들에서 함께 인용되는 동시 인용

관계를 바탕으로 논문 간 동시 인용 네트워크를 구축하고, 역시 논문을 군집화 하여 세부 주제를 정의한다(Small, 1973). 그러나 동시출현단어분석은 키워드 수준에서 명확하게 군집을 구분하고 정의하기 어렵다는 한계가 있으며, 동시인용분석은 개별 논문을 하나의 군집으로만 분류하게 되므로, 한 논문에 포함된 다양한 세부 주제를 반영할 수 없다는 문제가 있다(Ding and Chen, 2014).

이러한 한계점을 보완하기 위해 확률적 토픽 모델링 기법을 적용하면 논문에 포함된 연구 주제를 보다 면밀하게 분석할 수 있다(Yan, 2015). 토픽 모델링은 비정형 문서 집합에 잠재되어 있는 주제를 발견하기 위한 알고리즘으로 대량의 문서 집합에 적용되며 다양한 종류의 데이터에 적용 가능하다(Blei, 2012). 토픽 모델링 알고리즘을 통해 산출된 주제별 키워드를 활용하여 방대한 양의 문서들에 내포되어 있는 의미 있는 주제를 추정할 수 있다. 이런 장점으로 최근 특정 학문 분야에 대한 다양한 주제를 식별함에 있어서 토픽 모델링 기법의 활용이 급격히 증가하고 있다.

이에 본 연구에서도 토픽 모델링 기법을 활용하여 국내 산업공학 분야의 연구 주제를 탐색한다. 본 연구는 대한산업공학회 발간 학술지에 2001년부터 2015년까지 게재된 1,242편의 논문의 초록 정보를 바탕으로, 토픽 모델링 기법 중 가장 대표적인 알고리즘인 LDA(Latent Dirichlet Allocation)를 적용하여 산업공학의 연구 주제를 식별한다. 도출된 50개의 주제들의 산업공학 세부 분야별 분포를 파악하고, 비중이 높은 연구 토픽들을 기간별로 산출한다. 또한 시간에 따른 토픽별 비중 변화 추이를 바탕으로 유망 및 쇠퇴 연구 토픽을 도출한다.

이후 본 논문의 구성은 다음과 같다. 제 2장에서는 토픽 모델링 기법을 활용한 연구 주제 분석에 대해 다루고, 제 3장에서는 본 연구에 사용된 방법에 대해서 설명한다. 제 4장에서는 토픽 모델링을 통해 도출된 산업공학 연구 토픽을 정의하고 분포를 파악한다. 제 5장에서는 기간별 토픽들의 비중 변화를 바탕으로 산업공학 분야의 핵심 토픽을 추출한다. 마지막 장에서는 결론 및 향후 연구 방향에 대해 기술한다.

2. 토픽 모델링을 이용한 연구 주제 분석

텍스트 마이닝은 비정형 데이터인 텍스트로부터 유의미한 정보를 추출하는 기법으로, 다양한 분야에서 텍스트 마이닝 기법을 활용한 연구가 활발하게 진행되고 있다(Cho *et al.*, 2012; Go *et al.*, 2011; Kim *et al.*, 2012). 텍스트 마이닝에서 분석의 대상이 되는 텍스트 자료는 용어-문서 행렬(term-document matrix)의 형태로 가공되며 이것은 텍스트 데이터와 해당 자료 내에서 사용된 용어들 간의 관계를 정형적인 수치형 형태의 행렬로 나타내는 기법이다. 텍스트 데이터 집합을 이용해 구축하는 용어-문서 행렬은 일반적으로 0의 값이 아주 많은 희소행렬(sparse matrix)의 성질을 가진다. 이러한 용어의 형태만을 이

용하여 정형적 자료 분석 기법으로 텍스트 데이터를 해석하게 되면 동의어, 유의어, 동음다의어에 대한 구별이 제대로 이루어지지 않고 편향된 결과가 도출될 수 있다(Kim, 2016). 따라서 텍스트 마이닝 분석을 통해 의미 있는 결과를 도출하기 위해서는 용어의 모양이 아니라 실제 그 용어들이 가지고 있는 의미를 해석할 필요가 있으며 이에 대한 대안 중 하나로 토픽 모델링을 적용할 수 있다.

텍스트 마이닝 기법 중 하나인 토픽 모델링은 비정형 텍스트에서 주제를 찾아 각 주제 간의 연관 관계와 시간에 따른 변화를 분석해 주며, 구조화되지 않은 방대한 문서의 집합에서 핵심 주제를 탐색하는 통계 알고리즘이다(Blei, 2012). 토픽 모델링은 각 문서는 주제들의 조합이며, 각 주제는 단어들을 통해 확률 분포를 갖는다는 것을 가정한다. 따라서 문서에서 빈번하게 사용되는 단어들을 관찰함으로써 문서 중의 주제와 이런 주제들이 문서에서 어떤 분포인지 추정할 수 있다(Blei, 2012).

방대한 문헌 데이터에서 의미 있는 주제를 찾아내기 위해 LSI(Latent Semantic Indexing), pLSI(probabilistic Latent Semantic Indexing), LDA 등의 토픽 모델링 알고리즘을 활용할 수 있

다. LDA는 현재 모델을 업데이트할 필요 없이 새로운 문서의 주제들을 추정할 수 있다. 또한 문서 크기와 상관없이 매개 변수들(parameters)의 수가 고정되어 있어 방대한 양의 문서를 처리할 수 있다(Blei et al., 2003). 따라서 LDA는 지속적으로 생성되는 큰 텍스트 문서 데이터를 분석에 쉽게 적용할 수 있기 때문에 과적합(overfitting) 문제를 피할 수 있다.

LDA의 기본적인 아이디어는 문서를 구성하는 각 단어들이 주제 집합에 대한 확률 분포를 가지고 있으며 이 분포에 따라 각 문서는 특정 주제들로 구성되고, 주제들에 포함된 단어들이 각 문서를 구성하게 된다는 것이다. <Figure 1>은 LDA의 문서 생성 과정을 시각화한 것이다.

관찰된 변수인 단어 $w_{d,i}$ 와 문서별 주제 분포 매개 변수 α , 주제별 단어 분포 매개 변수 β 를 통해 보이지 않는 변수인 문서별 주제 비율(Per-document topic proportions, θ_d), 단어별 주제 지정(Per-word topic assignment, $z_{d,i}$), 주제별 단어의 확률 분포(Per-topic word distributions, ϕ_k)를 예측한다. α 와 β 는 단어를 추출하기 위해 사용된 모수이며 전체 문서에 동일한 값을 가진다. α 는 문서별 주제 비율인 θ_d 를 결정하는 매개 변수이며 Dirichlet 분포를 따르는 θ_d 는 α 에 따라 Dirichlet 분포 형

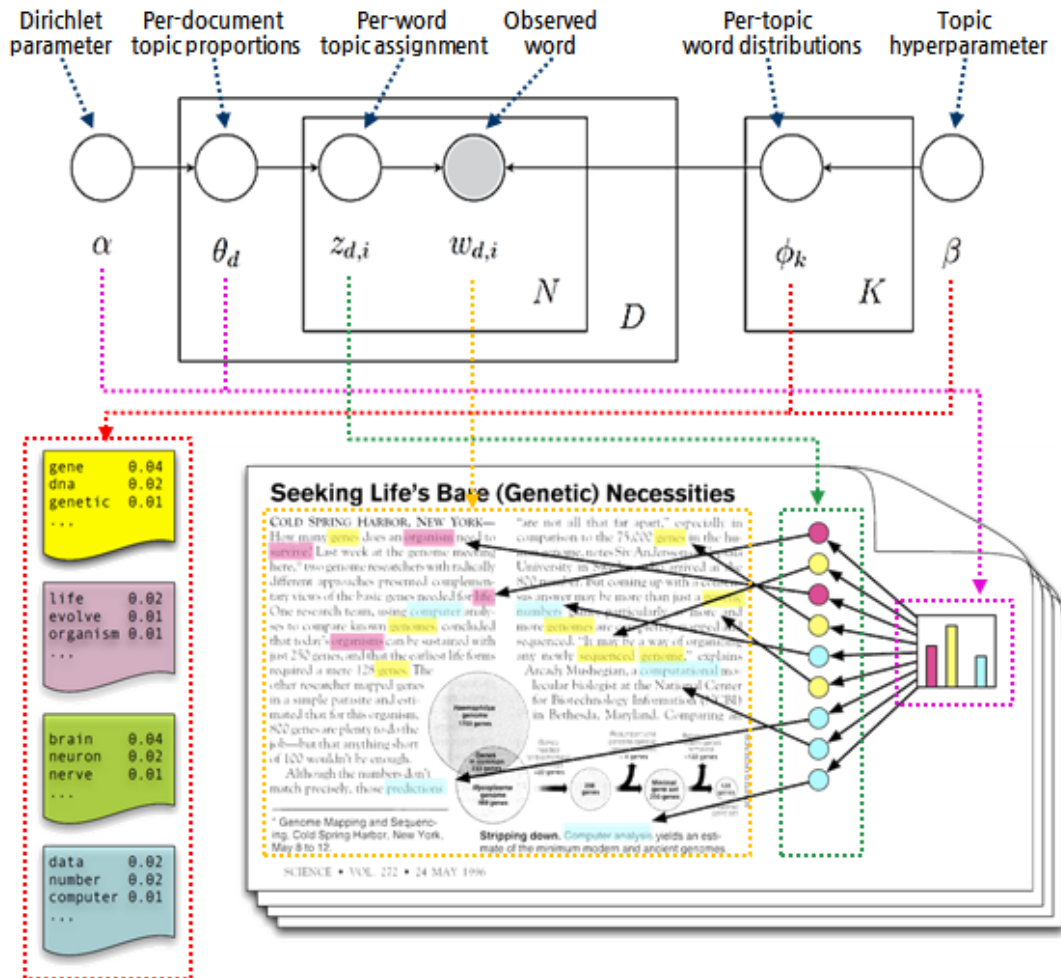


Figure 1. Document Generation Process of LDA(Blei et al., 2003)

태가 결정된다. 그리고 β 도 주제별 단어의 확률 분포인 ϕ_k 를 결정하는 매개 변수로 ϕ_k 는 β 에 따라 Dirichlet 분포 형태가 결정된다. 따라서 문서별 주제 비율인 θ_d 에 따라 문서 내에 존재하는 단어들의 주제인 $z_{d,i}$ 가 결정되며, $z_{d,i}$ 와 각 단어에 대한 전체 주제에 대한 비율인 ϕ_k 에 따라 단어 $w_{d,i}$ 가 결정된다.

LDA 알고리즘을 통해 대량의 논문 데이터에서 문서 집합의 주제들과 각 문서별 주제 비율, 각 단어들이 각 주제에 포함될 확률들을 알아낼 수 있으며, 추출된 주제를 기반으로 연구 동향을 분석할 수 있다(Tsai, 2011). 이에 토픽 모델링이 특정 학문 분야에 적용된 사례는 문헌 정보학(Park and Song, 2013; Cho, 2011), 생물정보학(Song and Kim, 2012), 컴퓨터 공학(Mimno and McCallum, 2006), 정치학/심리학/계량서지학(Blei, 2012), 기록관리학(Lee et al., 2007), 정보시스템학(Ahn et al., 2015), 신재생에너지(Shin et al., 2015), 통계학(Kim, 2014), 바이오 의료학(Won, 2014), 데이터정보과학(Kim and Park, 2015)으로 다양한 분야에서 활용되고 있다. 이와 같이 토픽 모델링을 통한 다양한 학문 분야에 대한 다각적인 분석이 이루어졌으나 산업공학 자체에 대한 연구는 그동안 텍스트 마이닝 기법을 사용하여 출현빈도분석, 구간별 분석, 군집화, 연관성 분석을 통해 국내 산업공학 분야에서 사용되고 있는 다양한 연구 기법을 체계적으로 분석한 연구(Cho et al., 2014)와 다양한 데이터마이닝 방법으로 산업공학 분야의 대표적인 국제학술지인 IIE Transactions에 출판된 논문을 분석하여 산업공학 내 연구의 현황과 추이를 관찰한 연구(Cho & Kim, 2012)로 토픽 모델링을 통한 산업공학의 심층적인 연구 동향을 분석한 사례를 찾아보기 어렵다. 따라서 본 연구에서는 최근 연구 동향 및 학문의 지적 구조 분석에 많이 사용되고 있는 토픽 모델링을 적용하여 국내 산업공학 분야에 대한 전반적인 연구 동향을 분석하고자 한다.

3. 연구 방법

국내 산업공학 분야 연구 동향 분석을 위해 대한산업공학회가 발간하는 대한산업공학회지와 IE Interfaces를 선택하였다. 두 학술지에 2001년부터 2015년도까지 게재된 논문 1242편을 수집하였으며, 구체적으로 대한산업공학회지 총 61권 638편의 논문과 IE Interfaces 총 48권 604편의 논문이 수집되었다. IE Interfaces의 경우 2013년 2월부터 대한산업공학회지와 통합되었기 때문에 이후로는 대한산업공학회지 논문만 수집되었다. 논문 데이터는 한국교육학술정보원(KERIS)이 제공하는 학술연구정보서비스(RISS)를 통해 수집하였다.

수집된 논문의 텍스트 데이터 중 영문 제목 및 초록 부분만을 추출하여 분석에 활용하였다. 일부 연구에서 논문의 전체 텍스트를 사용하여 LDA 기반의 연구 동향 분석이 시도되기는 했지만, 상대적으로 분량이 많지 않은 논문의 초록과 같은 텍스트 데이터를 사용한 LDA 기반의 분석보다 장점을 가지고

있다는 것은 확인되지 않았다(Song et al., 2015; Song and Kim, 2013). 또한 대다수의 연구 논문들이 관련되어 있는 기존 연구를 소개하고, 그 의미와 한계를 설명하는 문헌 검토 부분을 포함하기 때문에 대상 논문 데이터의 주요 의미 및 관련된 연구 방법, 연구 결과와 다른 방향으로 LDA 모형이 만들어질 수 있다. 따라서 본 연구에서는 1,242편 논문의 영문 제목 및 초록 데이터를 바탕으로 LDA를 수행하였다.

LDA 수행에 앞서 몇 가지 전처리 과정을 수행하다. 우선 분석에 필요한 단위로 텍스트를 나누는 tokenization을 수행한 뒤 주어를 포함한 a, an, the 등의 관사나 of, by, form 등의 전치사 등 분석에 불필요한 불용어(stopwords)를 모두 제거했다. 일반적인 불용어와 함께, 논문 초록에 빈번하게 출현하는 “study”, “paper”, “result” 등의 명사와 “analyze”, “examine”, “present”, “propose”, “discuss” 등의 단어를 추가로 불용어로 처리하였다. 다음으로 문장 내에서 다양한 형태로 사용되는 단어들의 표제어(lemma)를 추출하는 lemmatization 과정을 통해 각 단어들의 표제어를 추출했다. 자연어에서 단어는 문장에서 다른 형태로 사용될 수 있다. 예를 들어, “simulate”는 동사형인 “simulated”, “simulating”, “simulates”와 명사형 “simulation”, “simulations” 또는 형용사형 “simulative”, 부사형 “simulatively”으로 사용될 수 있다. 이때 단어의 어근을 찾는 대표적인 방법은 stemming과 lemmatization 두 가지가 있다. 두 방법의 가장 큰 차이점은 stemming은 단어 그 자체만을 고려하지만 lemmatization은 그 단어가 문장 내에서 어떤 품사로 사용되었는지가 고려된다. 따라서 “simulate”와 관련된 위의 일곱 개의 단어는 stemming을 통해 모두 “simul” 단어로 분석되는데 반해 lemmatization은 “simulated”, “simulating”, “simulates”는 “simulate”로 “simulation”, “simulations”는 “simulation”으로 “simulative”, “simulatively”는 “simulative”로 추출한다. 즉, stemming 방법은 lemmatization에 비해 차원(dimensionality)을 줄일 수 있지만 품사 정보가 손실되고, 해석하기 어려운 단어 형태가 될 수 있다. 주제를 정의하기 위한 단어 분포의 해석이 본 연구의 주요 목표 중 하나이기 때문에 lemmatization 방법이 stemming보다 더 적합하다.

영문 제목 및 초록 부분에서 의미적 정보가 포함된 단어 집합을 추출한 뒤 R 패키지 “topicmodels”를 수정 및 활용하여 분석을 수행하였다(Hornik and Grün, 2011). 추론 방법으로는 깃스 샘플링(Gibbs sampling) 알고리즘을 채택하였고, 반복횟수는 1,000, 5,000, 10,000, 50,000으로 다양하게 적용하였으며, 토픽의 개수를 20, 30, 50, 100으로 변화시켜 실험을 수행하였다. Dirichlet 파라미터인 α 는 문서 집합에 의해 정해진 기본 값을 사용하였고, β 는 패키지의 기본 값 0.1을 사용했다. 실험 결과를 바탕으로 가장 적합한 토픽 개수 및 반복 횟수를 선정하기 위한 지표(metrics)들이 존재하지는 않지만, 일반적으로 해석이 용이하고 더 의미 있다고 판단되는 결과를 연구자가 직접 판단하는 것이 바람직하다(Song et al., 2015; Song and Kim, 2013). 본 연구에서는 최종적으로 반복 횟수 10,000회에 토픽 수 50개 실험 결과를 채택하였다.

Table 1. 50 Topics of Industrial Engineering and Five Most Relevant Words

[T1] Product design & development	[T2] Assembly line	[T3] Miscellaneous 1	[T4] Scheduling	[T5] Diffusion modeling	[T6] Data Mining - Supervised Learning	[T7] Optimization	[T8] Performance measurement	[T9] Service management	[T10] Empirical case study
product design development material	assemble line virtual manufacture work	method performance model optimization cycle	schedule job machine minimize schedule	estimate distribution parameter pattern diffusion	datum classification prediction learn predict	optimal determine present total numerical	performance quality measure improve evaluate	service environment provide increase framework	analysis study analyze relationship case
[T11] Reliability engineering	[T12] Task analysis	[T13] Business process management	[T14] Quality inspection	[T15] RF identification and security	[T16] Experimental design	[T17] Mathematical modelling	[T18] Database management	[T19] Queuing management	[T20] Biomechanics
cost maintenance reliability failure consider	size type field significant optimum	process business evaluation bpm collaborative	item component procedure facility inspection	rfid risk technology safety security	factor design effect condition experiment	model propose important mathematical recent	datum support database base analysis	time queue dynamic wait delay	posture effect motion worker human
[T21] Information system management	[T22] Data Mining - Unsupervised Learning	[T23] Statistical quality control	[T24] Demand forecasting & inventory management	[T25] Cognitive ergonomics	[T26] Supply chain management (SCM)	[T27] Semiconductor manufacturing	[T28] Mobile UI design	[T29] Fault detection and monitoring	[T30] Group technology
system information integrate implement realtime	propose rule technique cluster tree	control average sample chart state	demand policy inventory forecast order	workload affect investigate experiment cognitive	production supply plan chain present	manufacture semiconductor order yield complex	interface human layout mobile design	base monitor collect individual detection	part group machine propose approach
[T31] Warehouse design	[T32] Miscellaneous 2	[T33] Software engineering	[T34] Physical ergonomics	[T35] Technology management	[T36] Machining	[T37] Financial engineering	[T38] Multiple criteria decision making	[T39] Strategic management	[T40] Simulation
operation block warehouse container	present study reduce perform measurement	develop project development management software	hand korean significant age characteristic	technology industry research activity strategy	tool machine operation tolerance operational	market investment price strategy index	decision make propose criterion multiple	company business collaboration management sigma	simulation result develop effectiveness methodology
[T41] Network traffic management	[T42] Heuristic optimization	[T43] Energy system engineering	[T44] Logistics planning	[T45] Optimization in telecommunication networks	[T46] Customer satisfaction	[T47] Efficiency analysis	[T48] Optimization for wireless networks	[T49] Vehicle routing	[T50] Literature review
network traffic allocation capacity efficient	problem algorithm solution heuristic solve	power module program equipment operator	logistic transportation plan develop delivery	location area call scheme propose	customer customer satisfaction search	efficiency evaluation analysis efficient evaluate	function level node wireless performance	vehicle suggest route travel route	research issue introduce discuss review

4. 산업공학 연구 주제 현황

LDA 결과 도출된 토픽별 단어 분포를 바탕으로 50개 토픽의 명칭을 정의하였다. 토픽의 명칭은 각 토픽별로 출현 확률이 높은 단어들을 바탕으로 정의하게 되는데, 단어(word) 자체만으로는 특정 의미를 지니고 있지 않는 경우가 있기 때문에 명확한 정의가 어려울 수 있다. 이에 본 연구에서는 각 토픽이 높은 비중을 차지하고 있는 상위 10개 논문을 추출하여, 이들의 제목과 초록을 참조하여 50개의 토픽 명칭을 정의하였다. 토픽 내 단어들의 연관성이 적거나 특정 주제로 지정하기 힘든 두 토픽은 Miscellaneous로 처리하였다. 정의된 토픽은 아래 <Table 1>에 나타나 있으며, 각 토픽별로 출현 확률이 높은 상위 5개 단어가 제시되어 있다. <Table 1>에 나타난 50개의 토픽은 2000년대 국내 산업공학 연구에서 다루어진 세부 연구 주제들이다.

도출된 50개의 토픽들을 산업공학의 세부 분야별로 분류하여 국내 산업공학 연구 주제의 분포를 분석하였다. 산업공학 세부 분야 분류는 대한산업공학회(2006)의 보고서에서 활용한 분류 체계를 활용하였다. 동 분류 체계는 산업공학을 시스템 분석(system analysis), 생산 및 물류(production/logistics), 품질(quality), 인간공학(ergonomics), 정보시스템(information system), 경영공학(management engineering), 기초과학(basic science)의 7개 세부 분야로 나누고 있으며, 이 중 기초과학을 제외한 6개의 분야에 도출된 토픽들을 할당하였다. 6개 분야에 속하지 않는 토픽은 기타(etc)로 분류하였다. 50개 토픽 별 분류 결과는 <Appendix A>에 제시되어 있다. <Table 2>는 세부 분야별 토픽의 수를 나타낸 것이며, 이를 <Figure 2>로 도식화하였다.

가장 많은 토픽을 포함하는 분야는 시스템 분석(13)과 생산 및 물류(13) 분야로 나타났으며, 이 두 분야가 전체의 절반 이상을 차지하고 있다. 반면 나머지 네 개 분야들은 4~6개의 토픽을 포함하고 있는 것으로 나타났다. 최근 다양한 분야와의 융합을 통해 산업공학의 외연이 확대되었지만, 산업공학의 본질이라고 할 수 있는 시스템 분석 및 최적화 기법 관련 연구가 여전히 핵심적인 연구 주제로 다루어지고 있으며, 전통적인 생산 및 물류 분야 역시 확고한 위치를 차지하고 있음을 알 수 있다. 반면 품질과 인간공학 분야는 두 분야에 비해 상대적으

Table 2. Number of Topics for Six IE Sub-Categories

Mark	Category	Number of topic
S	system analysis	13
P	production & logistics	13
Q	quality	4
E	ergonomics	5
I	information system	6
M	management engineering	5
N/A	etc	4

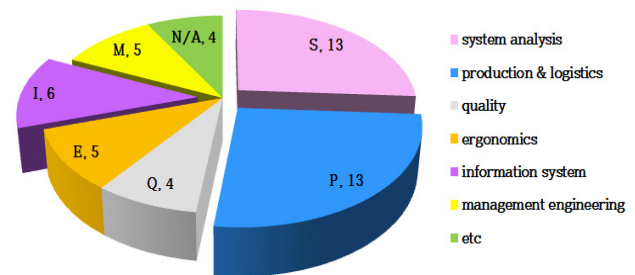


Figure 2. Topic Distribution Over Six IE Sub-Categories

로 낮은 비중을 차지하고 있다. 비교적 최근에 산업공학의 범주에 포함된 정보시스템과 경영공학 분야도 일정한 비중을 차지하고 있으나, 토픽 개수 관점에서는 아직 산업공학의 주류(mainstream)라고 보기에는 미흡한 수준인 것으로 나타났다.

5. 산업공학 연구 주제 트렌드

5.1 기간별 핵심 연구 토픽

산업공학 분야의 핵심 연구 토픽을 도출하기 위해 2001년에서 2015년 동안의 연구 주제 점유율을 분석하였다. 먼저 <Table 3>은 15년의 전체 기간에 대한 상위 10개 연구 토픽을 나타낸 것이며, 50개의 모든 토픽의 점유율은 <Appendix A>에 나타나 있다. [T42] Heuristic optimization이 가장 비중이 높은 토픽으로 나타났으며, 뒤를 이어 [T35] Technology management, [T21]

Table 3. Top 10 Topics in IE Research

Rank	Topic	Name	Category	Share
1	Topic 42	Heuristic optimization	S	3.429%
2	Topic 35	Technology management	M	2.598%
3	Topic 21	Information system management	I	2.290%
4	Topic 17	Mathematical modelling	S	2.201%
5	Topic 39	Strategic management	M	2.196%
6	Topic 6	Data Mining-Supervised Learning	I	2.193%
7	Topic 37	Financial engineering	M	2.172%
8	Topic 3	Miscellaneous 1	N/A	2.155%
9	Topic 13	Business Process Management(BPM)	I	2.146%
10	Topic 11	Reliability engineering	Q	2.123%

Information system management, [T17] Mathematical modelling, [T39] Strategic management 등이 높은 순위를 차지하였다. 특이할 만한 사항으로는 분야별 토픽 개수 측면에서는 비중이 낮게 나타난 경영공학과 정보시스템 분야의 토픽들이 상위 10개 토픽 안에 각각 3개나 포함되어 있다는 사실이다. 즉 상대적으로 신생 분야라고 할 수 있는 이들 두 분야는 전체 산업공학 분야에서 차지하는 비중은 아직 높지 않지만, 개별 토픽 관점에서는 매우 활발한 연구가 진행되고 있음을 알 수 있다.

시간에 따른 핵심 연구 토픽의 변화 추이를 관찰하기 위해 분석 대상 기간 15년을 5년 단위로 각각 세 기간으로 구분한 뒤 기간별 핵심 연구 토픽 상위 10개를 <Table 4>와 같이 도출하였다. <Table 4>를 통해 시간 흐름에 따른 핵심 토픽의 변화가 관찰된다. 세 기간에 공통적으로 나타나는 토픽은 [T42] Heuristic optimization, [T21] Information system management의 두 개 밖에 없으며, 이는 시간에 따른 산업공학의 핵심 연구 주제의 변화가 상당히 급속하게 진행되고 있음을 의미한다. [T42] Heuristic optimization의 경우 세 기간 모두에서 상위에 위치하지만 [T21] Information system management는 점유율이 점점 줄어드는 것을 볼 수 있다. Period 1에 나타난 주제들 중 [T20] Biomechanics, [T2] Assembly line, [T1] Product design & development, [T30] Group technology, [T26] Supply chain management(SCM)의 5개 주제는 Period 2부터는 순위권에 포함되어 있지 않다. 반면 Period 3의 [T37] Financial engineering, [T25] Cognitive ergonomics, [T47] Efficiency analysis, [T10] Empirical case study 4개의 주제는 Period 1과 Period 2에서 나타나지 않던 주제로, 최근 들어 상위 주제로 포함된 토픽들이다. 특이할 만한 사항은, [T13] Business Process Management(BPM)가 Period 1에서 Period 2로 기간이 변경되면서 점유율의 순위가 급상승된 뒤에 다시 Period 3에서는 상위 주제에서 제외되었고, [T35] Technology management는 Period 2에서 Period 3으로 기간이 변경되면서 가장 높은 점유율을 가진 주제로 위상이 증가하였다. 이와 유사하게 [T6] Data mining-Supervised learning도 Period 2에서 Period 3으로 오면서 점유율이 급격히 상승하였음을 알 수 있다.

Table 4. Top 10 topics of each of the three periods

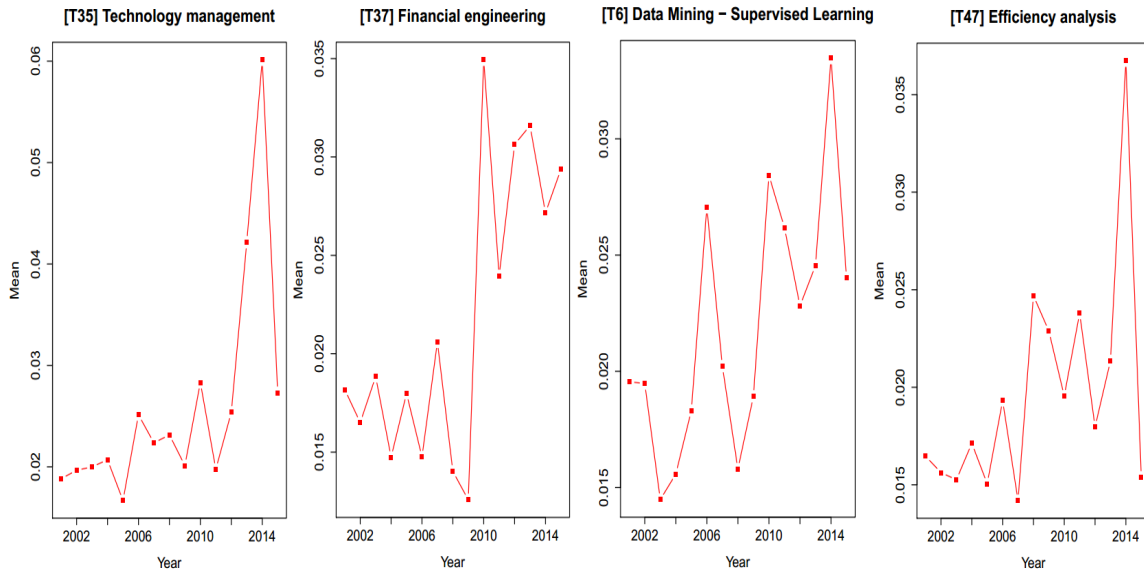
Rank	Period 1(2001~2005)	Period 2(2006~2010)	Period 3(2011~2015)
1	[T42] Heuristic optimization	[T42] Heuristic optimization	[T35] Technology management
2	[T20] Biomechanics	[T39] Strategic management	[T37] Financial engineering
3	[T2] Assembly line	[T13] Business process management(BPM)	[T42] Heuristic optimization
4	[T21] Information system management	[T35] Technology management	[T6] Data mining-Supervised learning
5	[T17] Mathematical modelling	[T21] Information system management	[T25] Cognitive ergonomics
6	[T4] Scheduling	[T7] Optimization	[T47] Efficiency analysis
7	[T1] Product design & development	[T4] Scheduling	[T10] Empirical case study
8	[T30] Group technology	[T11] Reliability engineering	[T3] Miscellaneous 1
9	[T26] Supply chain management(SCM)	[T6] Data mining-Supervised learning	[T17] Mathematical modelling
10	[T13] Business process management(BPM)	[T9] Service management	[T21] Information system management

5.2 유망 및 쇠퇴 연구 토픽

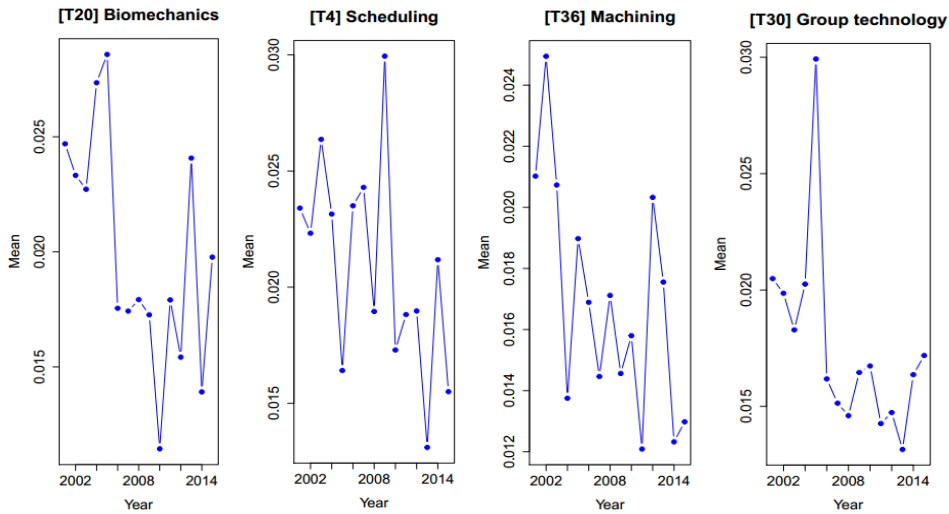
제 5.1절에서 시간 흐름에 따른 연구 토픽의 점유율 변화 분석을 통해, 기간별 핵심 연구 주제와 함께 연구자들의 관심이 낮아지는 연구 주제와 새롭게 주목을 받는 연구 주제 등의 전반적인 변화를 관찰할 수 있다. 하지만 이것은 5년 단위로 구분된 점유율에 대한 분석이므로 매년 변화되는 점유율을 반영한 것이 아니며, 상위 10개의 주제로만 분석 대상을 한정하였기 때문에 개별 주제의 연도별 상승과 하락 추세를 정확하게 파악하는데 한계가 있다. 따라서 50개 주제의 15년 기간 동안의 연도별 점유율 변화 추이를 바탕으로, 연구 주제의 점유율이 상승 추세를 보이는 유망(hot) 토픽과 감소 추세를 보이는 쇠퇴(cold) 토픽을 도출하였다. 각 연도를 독립 변수로 하고, 산업공학 주제의 연도별 점유율을 종속 변수로 하여 선형회귀 분석을 수행하였으며, 회귀 계수가 유의 수준 5%에서 유의한 양의 값을 가지면 유망 토픽으로, 음의 값을 가지면 쇠퇴 토픽으로 구분하였다 (Griffiths and Steyvers, 2004). 도출된 유망 및 쇠퇴 토픽의 목록과 회귀 계수는 아래 <Table 5>와 같으며, 각각의 유망 및 쇠퇴 토픽의 연도별 비중 변화 추이를 <Figure 3>에 나타내었다.

Table 5. Hot and cold topics in IE research

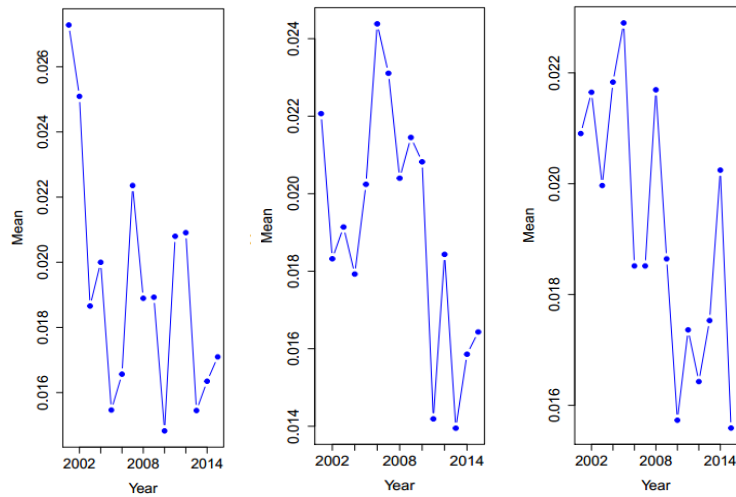
Topic	Category	Slope (×1000)	Hot/Cold
[T35] Technology management	M	15.8952	Hot
[T37] Financial engineering	M	11.4238	Hot
[T6] Data Mining-Supervised Learning	I	7.8452	Hot
[T47] Efficiency analysis	S	6.7253	Hot
[T20] Biomechanics	E	-6.2948	Cold
[T4] Scheduling	P	-5.1673	Cold
[T36] Machining	P	-5.1539	Cold
[T30] Group technology	P	-4.8718	Cold
[T19] Queuing management	S	-4.5384	Cold
[T9] Service management	P	-3.7519	Cold
[T41] Network traffic management	S	-3.6256	Cold



(a) Hot topics



[T19] Queuing managemen [T9] Service management T41] Network traffic managm



(b) Cold topics

Figure 3. Proportion Trends of Hot and Cold Topics in IE Research

앞서 Period 3에서 가장 높은 점유율을 차지한 [T35] Technology management가 비중이 가장 급격하게 증가한 토픽으로 나타났다. 최근 국내 일부 대학의 산업공학과를 중심으로 기술경영 대학원 운영이 활발히 이루어짐에 따라 기술경영 관련 연구가 산업공학 분야에서 급격히 증가한 것으로 보인다. [T37] Financial engineering 역시 급속한 증가 추세를 보이고 있으며, 이는 최근 주목 받고 있는 금융공학 분야가 산업공학 내에서 핵심적인 연구 분야로 발돋움하고 있음을 의미한다. 또한 [T6] Data mining-Supervised learning 토픽이 유망 토픽으로 선정된 것은 산업공학 분야가 기계학습의 지도학습 기법을 중심으로 최근 급증하고 있는 빅데이터 및 인공지능 관련 연구 수요에 대응하고 있음을 알 수 있다. 효율성 분석은 산업공학의 전통적인 주제임에도 불구하고, 최근 들어 자료포락분석 (data envelopment analysis)의 응용 연구가 활발히 이루어짐에 따라 [T47] Efficiency analysis가 유망 토픽으로 선정되었다. 반면, [T4] Scheduling, [T36] Machining, [T30] Group technology 등 전통적인 제조업 관련 연구 토픽은 지속적인 하락세를 보이고 있다.

6. 결론

본 연구는 국내 산업공학 분야의 연구 주제를 파악하기 위해 대한산업공학회지와 IE Interfaces에 2001년에서 2015년까지 게재된 1242개의 논문 영문 초록 데이터를 바탕으로 LDA 분석을 수행하였다. 이를 통해 국내 산업공학 연구의 50개 토픽을 도출하였으며, 산업공학의 세부 분야별 분포를 분석하였다. 또한 기간별 핵심 토픽 10개를 도출하고, 연도별 점유율 변화 추이를 바탕으로 유망 및 쇠퇴 토픽을 파악하였다.

지난 40년간 국내 산업공학 분야는 사회의 새로운 수요에 대응하기 위해 확장과 변화를 거듭해 왔으며, 향후 지속적인 발전 방향 정립을 위해서는 현재까지의 연구 추세와 현황에 대한 면밀한 분석이 선행되어야 한다. 본 연구는 최근 15년간 수행된 산업공학 연구의 핵심 주제들을 파악하고 이들의 변화 추세를 분석함으로써, 향후 산업공학 분야의 유망 연구 주제 설정 및 발전 방향 수립에 활용될 수 있는 구체적인 시사점을 제공했다는 측면에서 의의를 가진다. 산업공학 연구자들은 현재 수행 중인 연구 주제가 유망 또는 쇠퇴 분야인지를 파악할 수 있으며, 산업공학 내에서 해당 연구 주제의 위상을 가늠할 수 있다. 또한 학술지 편집위원들은 학술지에 게재된 논문들의 주제 및 성격이 편집 방침과 일치하는지를 확인할 수 있으며, 향후 편집 방침에 대한 비전 및 방향 설정에 활용할 수 있다. 연구 지원 기관에서는 도출된 유망 분야를 활용하여 산업공학 및 유관 분야의 유망 연구 과제 발굴이 가능하다.

그러나 본 연구는 대한산업공학회가 발간하는 국내 학술지 논문으로만 분석 범위를 제한하였는 한계가 있다. 한국경영과학회, 한국경영공학회, 한국산업경영시스템학회 등 산업공학 관련 학회에서 발간하는 학술지 논문으로 분석 범위를 확장할

필요가 있다. 또한 분석 범위를 국제 학술지 논문으로 확대하여 동일한 분석을 수행한다면, 산업공학 연구의 국제 동향 파악이 가능할 것이며, 이를 국내 결과와 비교함으로써 다양한 시사점을 도출할 수 있을 것이다. 또한 본 연구에서는 분석 기간인 15년 동안 게재된 모든 논문을 하나의 문서 집합으로 묶어서 LDA를 수행하였기 때문에, 연구 토픽의 세부적인 진화 과정을 파악할 수 없었다. 추후 연구에서는 각 기간별로 문서 집합을 구분하여 기간별로 서로 다른 토픽을 추출하고, 토픽 간의 관계를 정량적으로 측정 및 시각화하여 산업공학 분야의 연구 토픽의 진화 과정을 분석하는 것이 필요하다.

참고문헌

- Ahn, J. K., Lee, K. H., and Kim, H. W. (2015), Trend analysis of changes in information systems research : modeling and network analysis topics, *The Korea Society of Management Information Systems*, 561-570.
- Blei, D. M. (2012), Probabilistic Topic Models, *Communications of the Acm*, 55(4), 77-84.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003), Latent dirichlet allocation, *The Journal of machine Learning research*, 3, 993-1022.
- Callon, M., Courtial, J. P., Turner, W. A., and Bauin, S. (1983), From translations to problematic networks-An introduction to co-word analysis, *Social Science Information Sur Les Sciences Sociales*, 22(2), 191-235.
- Choi, B. K., Han, K. H., Jun, C. S., Lee, C. S., and Park, S. C. (2014), Review and Perspectives on the Research and Industrial Applications of Manufacturing Systems Engineering in Korea for 40 Years, *Journal of the Korean Institute of Industrial Engineers*, 40(6), 555-567.
- Cho, J. (2011), A study for research area of library and information science by network text analysis, *Journal of the Korean Society for Information Management*, 28(4), 65-83.
- Cho, G. H., Lim, S. Y., and Hur S. (2014), An Analysis of the Research Methodologies and Techniques in the Industrial Engineering Using Text Mining, *Journal of the Korean Institute of Industrial Engineers*, 40(1), 52-59.
- Cho, S. and Kim, S. (2012), Finding meaningful pattern of key words in IIE transactions using text mining, *Journal of the Korean Institute of Industrial Engineers*, 38(1), 67-73.
- Chung, M. K., Yun, M. Y., Park, J. H., Lee, I-S., and Lim, J-H. (2014), 40 Years of Ergonomics in Korea : Accomplishments, Challenges and 40 More Years Ahead, *Journal of the Korean Institute of Industrial Engineers*, 40(6), 568-579.
- Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., and Herrera, F. (2011), An approach for detecting, quantifying, and visualizing the evolution of a research field : A practical application to the Fuzzy Sets Theory field, *Journal of Informetrics*, 5(1), 146-166.
- Ding, W. and Chen, C. (2014), Dynamic topic detection and tracking : A comparison of HDP, C-word, and cocitation methods, *Journal of the Association for Information Science and Technology*, 65(10), 2084-2097.
- Go, G., Jung, W., Shin, Y., Park, S., and Jang, D. (2011), A study on development of patent information retrieval using textmining, *Journal of the Korea Academia-Industrial Cooperation Society*, 12(8), 3677-3688.
- Griffiths, T. L. and Steyvers, M. (2004), Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(1), 5228-5235.
- Hornik, K. and Grun, B. (2011), topicmodels : An R package for fitting topic models. *Journal of Statistical Software*, 40(13), 1548-7660.

- Jeong, B. K. and Lee, H. Y. (2016), Industrial Engineering as a Multidisciplinary Field : Exploring the Structure of Academic Convergence in Industrial Engineering by Journal Citation Network Analysis, *Journal of the Korean Institute of Industrial Engineers*, **42**(3), 182-197.
- Korean Institute of Industrial Engineers (2006), Domestic and international through case studies and the job analysis creative engineering education system and curriculum development : Focus on Industrial Engineering.
- Kim, G. H. and Park, C. L. (2015), Analysis of English abstracts in Journal of the Korean Data and Information Science Society using topic models and social network analysis, *Journal of the Korean Data and Information Science Society*, **26**(1), 151-159.
- Kim, H. B. (2016), Text Mining for Korean Using Topic Models Based on Latent Dirichlet Allocation Technique, *Graduate School of Dongguk University*, Seoul, 5-7.
- Kim, J. and Jeong, C. (2012), Analysis of trend in construction using text mining method, *Journal of The Korean Digital Architecture · Interior Association*, **12**(2), 53-60.
- Kim, Y. K. and Kim, J. Y. (2014), OR/MS in Korea : The Review and Outlook, *Journal of the Korean Institute of Industrial Engineers*, **40**(6), 592-608.
- Lee, J. Y., Moon, J. Y., and Kim, H. J. (2007), Examining the Intellectual Structure of Records Management and Archival Science in Korea with Text Mining, *Korean Society for Library and Information Science*, **41**(1), 345-372.
- Mimno, D. and McCallum, A. (2008), Topic models conditioned on arbitrary features with Dirichlet multinomial regression, *Proceedings of the 24th Conference in Uncertainty in Artificial Intelligence(UAI 2008)*, UAI-P-2008-PG-411-418.
- Park, J. H. and Song, M. (2013), A Study on the Research Trends in Library and Information Science in Korea using Topic Modeling, *Journal of the Korean Society for Information Management*, **30**(1), 7-32.
- Park, J. W., Shin, H. S., Kim, K-D., Jeong, H-I., and Lee, J. C. (2014), Production Planning and Control in Korea : with Emphasis on the Role of Industrial Engineers, *Journal of the Korean Institute of Industrial Engineers*, **40**(6), 580-591.
- Park, Y. T. (2015), *Industrial Engineering-systems approach*, Saeng Neung Press, Paju-si, Korea.
- Ramos-Rodriguez, A. R. and Ruiz-Navarro, J. (2004), Changes in the intellectual structure of strategic management research : A bibliometric study of the Strategic Management Journal, 1980-2000, *Strategic Management Journal*, **25**(10), 981-1004.
- Shin, K. S., Choi, H. R., and Lee, H. C. (2015), Topic Model Analysis of Research Trend on Renewable Energy, *Journal of the Korea Academia-Industrial cooperation Society*, **16**(9), 6411-6418.
- Salvendy, G. (2001), *Handbook of Industrial Engineering : Technology and Operations Management*. Wiley-Interscience, New York City, United States.
- Small, H. (1973), Co-citation in the scientific literature : A new measure of the relationship between two documents, *Essays of an Information Scientist*, **24**(4), 265-269.
- Song, M. and Kim, Y. S. (2012), Detecting the knowledge structure of bio-informatics by mining full-text collections, *Scientometrics*, **96**(1), 183-201.
- Song, M., Heo, G. E., and Lee, D. (2015), Identifying the landscape of Alzheimer's disease research with network and content analysis, *Scientometrics*, **102**(1), 905-927.
- Song, M. and Kim, S. Y. (2013), Detecting the knowledge structure of bio-informatics by mining full-text collections, *Scientometrics*, **96**(1), 183-201.
- Tsai, F. S. (2011), A tag-topic model for blog mining, *Expert Systems with Applications*, **38**(5), 5330-5335.
- Won, J. H. (2014), A study of the forecasting demand of medicine, *Information & Industrial Engineering Yunsei University*.
- Yan, E. (2015), Research dynamics, impact, and dissemination : A topic-level analysis, *Journal of the Association for Information Science and Technology*, **66**(11), 2357-2372.
- Yum, B. J., Seo, S. K., Yun, W. Y., and Byun J. H. (2014), Trends and Future Directions of Quality Control and Reliability Engineering, *Journal of the Korean Institute of Industrial Engineers*, **40**(6), 526-554.

<Appendix A> List of 50 Topics

Rank	Topic	Name	Category	Share
1	Topic 42	Heuristic optimization	S	3.429%
2	Topic 35	Technology management	M	2.598%
3	Topic 21	Information system management	I	2.290%
4	Topic 17	Mathematical modelling	S	2.201%
5	Topic 39	Strategic management	M	2.196%
6	Topic 6	Data mining-supervised learning	I	2.193%
7	Topic 37	Financial engineering	M	2.172%
8	Topic 3	Miscellaneous 1	N/A	2.155%
9	Topic 13	Business process management(BPM)	I	2.146%
10	Topic 11	Reliability engineering	Q	2.123%
11	Topic 25	Cognitive ergonomics	E	2.093%
12	Topic 7	Optimization	S	2.089%
13	Topic 4	Scheduling	P	2.088%
14	Topic 31	Warehouse design	P	2.073%
15	Topic 2	Assembly line	P	2.073%
16	Topic 24	Demand forecasting & inventory management	P	2.055%
17	Topic 1	Product design & development	P	2.041%
18	Topic 10	Empirical case study	N/A	2.016%
19	Topic 26	Supply chain management(SCM)	P	2.012%
20	Topic 34	Physical ergonomics	E	2.007%
21	Topic 20	Biomechanics	E	1.996%
22	Topic 47	Efficiency analysis	S	1.969%
23	Topic 16	Experimental design	Q	1.943%
24	Topic 27	Semiconductor manufacturing	P	1.943%
25	Topic 33	Software engineering	I	1.941%
26	Topic 5	Diffusion modeling	M	1.932%
27	Topic 19	Queuing management	S	1.924%
28	Topic 38	Multiple criteria decision making(MCDM)	S	1.924%
29	Topic 28	Mobile UI design	E	1.917%
30	Topic 41	Network traffic management	S	1.917%
31	Topic 9	Service management	P	1.912%
32	Topic 45	Optimization in telecommunication networks	S	1.905%
33	Topic 18	Database management	I	1.904%
34	Topic 50	Literature review	N/A	1.885%
35	Topic 46	Customer satisfaction	M	1.875%
36	Topic 44	Logistics planning	P	1.873%
37	Topic 49	Vehicle routing	P	1.871%
38	Topic 48	Optimization for wireless networks	S	1.863%
39	Topic 8	Performance measurement	S	1.856%
40	Topic 14	Quality inspection	Q	1.834%
41	Topic 40	Simulation	S	1.828%
42	Topic 12	Task analysis	E	1.827%
43	Topic 15	RF identification and security	P	1.827%
44	Topic 43	Energy system engineering	S	1.812%
45	Topic 23	Statistical quality control	Q	1.787%
46	Topic 32	Miscellaneous 2	N/A	1.766%
47	Topic 30	Group technology	P	1.757%
48	Topic 22	Data mining-Unsupervised learning	I	1.754%
49	Topic 29	Fault detection and monitoring	S	1.720%
50	Topic 36	Machining	P	1.690%