

동시단어분석을 이용한 품질경영분야 지식구조 분석

박만희*

부산가톨릭대학교 경영학과

The Analysis of Knowledge Structure using Co-word Method in Quality Management Field

Man-Hee Park*

Department of Business Administration, Catholic University of Pusan

ABSTRACT

Purpose: This study was designed to analyze the behavioral change of knowledge structures and the trends of research topics in the quality management field.

Methods: The network structure and knowledge structure of the words were visualized in map form using co-word analysis, cluster analysis and strategic diagram.

Results: Summarizing the research results obtained in this study are as follows. First, the word network derived from co-occurrence matrix had 106 nodes and 5,314 links and its density was analyzed to 0.95. Average betweenness centrality of word network was 2.37. In addition, average closeness centrality and average eigenvector centrality of word network were 0.01. Second, by applying optimal criteria of cluster decision and K-means algorithm to word co-occurrence matrix, 106 words were grouped into seven clusters such as standard & efficiency, product design, reliability, control chart, quality model, 6 sigma, and service quality.

Conclusion: According to the results of strategic diagram analysis over time, the traditional research topics of quality management field related to reliability, 6 sigma, control chart topics in the third quadrant were revealed to be declined for their study importance. Research topics related to product design and customer satisfaction were found to be an important research topic over analysis periods. Research topic related to management innovation was emerging state and the scope of research topics related to process model was extended to research topics with system performance. Research topic related to service quality located in the first quadrant was analyzed as the key research topic.

● Received 30 March 2016, 1st revised 3 May 2016, accepted 4 May 2016

†Corresponding Author(mhpark@cup.ac.kr)

© 2016, The Korean Society for Quality Management

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.

* 이 논문은 2014년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 연구되었음(NRF-2014S1A5A8014670)

Key Words: Co-word analysis, Text-mining, Social network analysis, Cluster analysis, Strategic diagram

1. 서 론

국제표준화기구 ISO에 따르면 품질경영이란 최고 경영자의 리더십 아래 품질을 경영의 최우선 과제로 하고 고객만족 확보를 통한 기업의 장기적인 성공은 물론 기업 구성원과 사회 전체의 이익에 기여하기 위해 경영활동 전반에 걸쳐 모든 구성원의 참가와 총체적 수단을 활용하는 전사적 종합적인 경영관리 체제라고 정의하고 있다. 품질경영은 품질방침 목표 및 책임을 결정하고 이들을 품질계획, 품질관리, 품질보증, 품질개선과 같은 도구를 통해 품질시스템 내부에서 실행되도록 전반적 관리기능에 관한 모든 활동을 말한다. 기업의 지속가능경영과 전략적 차원에서 품질경영의 중요성은 도요타의 대규모 리콜사태를 통해 확인할 수 있다. 도요타 자동차가 2009년부터 2010년 사이에 제작된 차종의 엑셀레이터 품질문제로 인한 집단소송과 관련, 최대 1,600만대의 차종에 새로운 안전장비를 부착하고 운전자 배상에 11억 달러를 지불하기로 합의했다는 보도가 있었다(Wall street journal, 2014). 품질경영의 중요성으로 인해 대부분의 기업이나 조직은 고객의 요구사항을 이해하고 고객요구사항의 변화행태와 유형분석을 통해 고객요구사항을 충족시키는 제품과 서비스를 제공하기 위해 모든 역량을 집중하고 있다. 품질이 기업이나 조직의 지속가능성을 보장하고 경쟁우위를 달성할 수 있는 핵심요소 가운데 하나로 받아들여지고 있다.

세계적인 정보 인프라 솔루션 기업 EMC가 정보기술 시장조사기관 IDC에 의뢰해 2011년에 발표한 ‘디지털 유니버스’ 보고서에 따르면 전 세계에서 새로 만들어지는 데이터 양이 2년마다 두 배씩 증가할 것이라고 전망하고 있다. IDC는 “올해 새로 생성될 디지털 정보량은 약 1.8ZB에 달할 것”이라며 예측하고 있으며, 보고서에 따르면 2009년에 생산된 전 세계 디지털 정보량은 0.8ZB였지만 2010년 1.2ZB, 2011년 1.8ZB로 급증하고 있다(중앙일보, 2011). 이런 폭발적인 데이터 양 증가로 인해 연구자들이 자신의 연구분야에서 주제영역을 발견하거나 주제영역들 간의 연관관계를 찾아내는 것이 점점 어려워지고 있으며, 정책입안자들도 과학의 역동성(Dynamics)을 연구전략 및 연구계획 수립으로 반영하는데 있어 어려움을 겪고 있다. 과학의 개념, 아이디어, 문제점들간의 관계를 매핑하기 위해 기존 방법론들은 상대적으로 적은 수의 전문가 견해를 바탕으로 이루어져 왔으며, 이런 접근방법이 사안의 중요성과 상황에 따라서 반드시 필요한 방법론이지만 몇 가지 약점을 가지고 있다고 지적하였다(He, 1999). 전문가 설문조사의 양이 적지 않다면 상당히 많은 비용이 발생하고, 전문가 설문조사의 양이 적다면 결과의 대표성에 문제제기가 있을 수 있다. 또한 과학이 발전해온 방식과 궤적 및 향후 발전방향에 대한 다양한 견해를 분석하는 문제는 매우 복잡할 수 있다는 것이다.

현재까지 수행된 품질경영분야 지식구조 분석 관련 연구의 대부분은 정성적인(Qualitative) 방법론을 적용하여 연구자의 주관적인 판단으로 자료의 신뢰성을 검증하고 연구결과를 바탕으로 지식구조 변화를 분석하여 미래 발전 방향을 제시하고 있다. 또한 품질경영 전문가 그룹이나 사용자 그룹을 대상으로 설문조사 분석을 수행하고, 통계분석 기법 적용을 통해 지식구조의 변화를 분석하여 이를 바탕으로 미래 지식구조의 발전행태를 예측하는 형태로 수행되어 왔다. 따라서 지식구조 탐색과 분석을 위해 개발된 정량적(Quantitative) 방법론과 이를 지원하기 위한 다양한 소프트웨어들을 활용하여 품질경영 분야의 지식구조를 분석하고 가시화를 통해 발전 행태를 살펴보고, 새롭게 부상하는 연구주제 및 시계열에 따른 차이점을 분석할 필요성이 있다. 따라서 본 연구는 동시단어분석을 이용하여 품질경영학회지의 연구문헌에 나타난 키워드들의 동시출현 관계와 연결강도 매핑을 통해 품질경영 분야의 발전추세와

핵심연구주제 및 새롭게 부상하고 있는 연구영역을 분석하고자 하였다. 또한 2002년부터 2014년까지 데이터를 2개 구간으로 분할하여 지식맵으로 가시화하고, 시계열에 따른 지식구조의 변화 형태를 분석하고 전략적 시사점을 도출하고자 하였다.

2. 선행연구 고찰

동시단어분석을 이용하여 지식구조 분석을 수행한 선행 연구들을 살펴보면 다음과 같다.

이정희(2011)는 지식은 체계화된 정보이므로 본질적으로 구조를 갖고 있으며 분과학문은 지식의 축적물로서 지식에 내재적으로 존재하는 구조를 바탕으로 성장하게 된다. 해당 연구에서는 예산이론 연구프로그램이 가지고 있는 지적 구조를 파악하고자 시도하였다. 기존의 연구가 예산연구자의 주관적 관점에 기반한 것에 비하여 내용분석과 연결망 분석을 통하여 양적인 분석을 시도하였다. 내용분석을 통하여 이론연구 프로그램의 연구의 증감, 개별이론 연구의 증감 등 연구인용패턴을 살펴보고, 문헌들의 연결망 분석으로 주요 이론군을 발견하고 연구프로그램들을 지도하는 이론을 식별하였다. 분석결과 예산이론연구가 지속적으로 증가하고 있고, 예산이론연구프로그램 내의 독자성을 가진 연구군으로 예산이론일반연구군, 단절균형이론연구군, 공공선택이론연구군이 병렬적으로 존재하였다. 예산이론가들이 관심을 갖는 주요이론 중 단절균형이론, 공공선택이론의 경우 이론적 성숙단계, 다중합리성이론은 이론적 발전단계, 거래비용이론은 이론적 정체상태에 놓여있다고 판단할 수 있었다.

김하진 외(2014)는 국내외 정보학 분야 학회지 동향을 파악하기 위하여 1990년부터 2013년까지 국내 정보관리학회지와 국외 JASIST의 논문 제목과 초록을 대상으로 텍스트 마이닝 기법을 이용하여 명사, 명사구 동시출현 분석을 수행하였다. 또한 분석대상 기간을 5개 구간으로 나누고 전체적인 동향을 살펴보기 위해 고빈도 동시출현단어 분석을 수행하였으며, 세부 주제 파악을 위해 상위 키워드와 동시에 출현하는 단어를 분석하였다. 분석결과 전반적으로 국내의 경우 정보관리학회지는 도서관, 정보 서비스, 정보 이용자, 정보 자료 조직 주제 분야가 많이 차지하고 있고, JASIST는 정보 검색, 정보 이용자, 웹 관련 정보학, 계량 정보학 중심으로 연구가 진행되고 있는 것으로 분석되었다.

이수경 외(2011)는 국내 간호학 연구의 지식구조를 파악하기 위해 1995년부터 2009년까지 15년간의 8개 간호학회지에 게재된 5,936편의 논문을 대상으로 사회 연결망 분석방법을 사용하여 지식구조를 분석하였다. 연구결과에 따르면 15년간 상위 1%를 차지하는 연구 주제는 간호중재, 삶의 질, 자기 효능, 자아 존중감, 건강증진, 사회적 지지, 노인, 자가 간호, 교육프로그램 등이 포함되었다. 또한 3년 단위로 구분하여 각 시기별로 새롭게 나타나는 연구 주제를 분석한 결과 최근 대두되는 노화불안, 정보역량, 다문화 가족 등의 개념을 통해 노인인구의 증가, 정보기술의 발전, 사회적 약자에 대한 책임감 증대 등의 사회적 변화와 간호학 지식 구조의 변화가 일치한다는 것을 분석하였다.

품질경영 분야 지식구조를 분석한 연구 중 가장 대표적인 국내연구로는 이돈희 외(2012)가 수행한 연구로 2007년부터 2011년까지 품질경영학회지 게재논문을 중심으로 품질연구의 발전을 고찰하였다. 해당 연구에서는 2007년부터 2011년까지 5개년 동안 게재된 연구논문 총 228편을 중심으로 연구분야의 주제, 연구방법, 산업군, 연구자의 소속 등을 기준으로 학회지 연구동향 분석을 통해 향후 연구방향의 시사점을 분석하였다. 해당 연구의 연구내용과 결과를 요약하면 다음과 같다. 첫째, 품질경영학회지에 게재된 연구주제를 학술대회 주제발표 항목 기준으로 분류하고, 교신저자의 소속기관을 기준으로 학계(전공분류)와 산업계(산업분류) 비율을 분석하였다. 둘째, 실증적 연구, 분석적 연구, 사례연구 등 연구방법론에 따른 분석을 수행하고, 통계적기법을 이용하여 연구 주제별, 학과별, 산업별

상호관계성의 유의성을 검증하였다. 셋째, 연구주제에 따른 게재 논문의 수는 연도별로 큰 변화 없이 관련 주제에 대한 연구가 진행되고 있는 것으로 분석되어 시대적 흐름을 크게 반영하고 있지 못한 것으로 판단하고, 시대적 요구 사항을 파악하여 스페셜 이슈(Special issue) 등을 통하여 학문적으로 시대적 흐름을 반영할 필요가 있다고 지적하였다.

타 연구분야에서 동시단어분석을 적용하여 연구주제 및 지식구조 변화 분석을 수행한 연구는 다양하게 존재하지만 품질경영분야에는 동시단어분석과 연구주제들간의 연관성분석을 통해 연구주제 영역의 특성이나 지식구조의 변화형태를 분석한 연구는 찾아보기 어려운 것으로 조사되었다.

3. 연구방법

본 연구는 품질경영분야 지식구조의 변화행태와 연구주제의 변화추세 분석을 위하여 동시단어분석과 군집분석, 지식맵 등의 기법을 이용하여 키워드의 네트워크 구조와 지식구조를 맵(Map) 형태로 가시화하고, 다양한 분석 지수를 이용하여 네트워크의 특성과 시간의 변화에 따른 구조변화를 분석하였다. 본 연구에서 수행한 연구절차와 연구내용을 정리하면 그림 1과 같다.

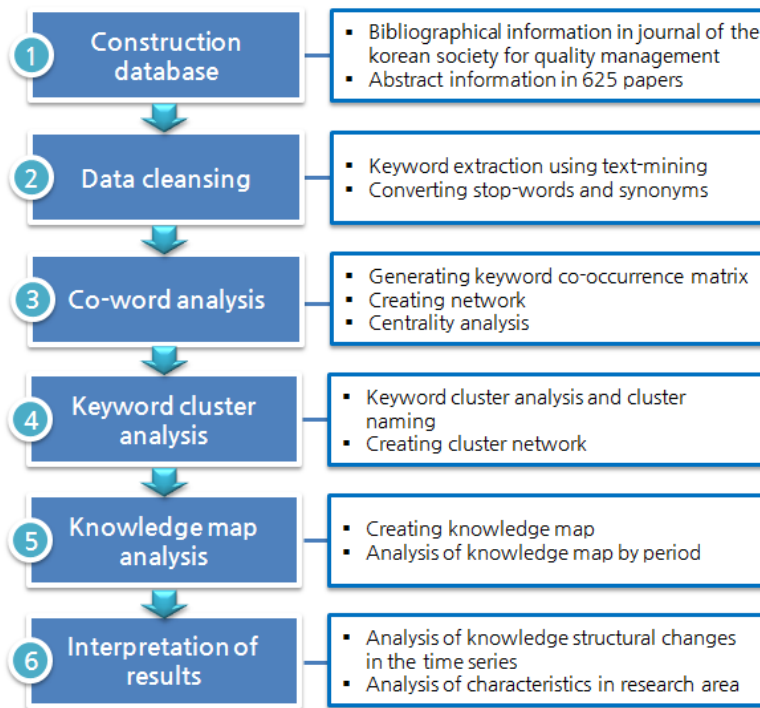


Figure 1. Research framework

본 연구에서 수행한 주요 연구절차와 핵심 연구내용 및 연구 방법론을 개략적으로 설명하면 다음과 같다. 데이터

베이스 구축단계에서는 2002년부터 2014년까지 품질경영학회지에 게재된 학술논문에는 포함된 영문제목, 영문초록, 발행연도, 저자, 소속 등의 서지정보를 이용하여 데이터베이스로 구축하였다. 한 논문의 영문제목, 영문초록, 발행연도, 저자, 소속 등의 데이터가 하나의 레코드로 구성되도록 하여 데이터베이스를 구축하였다. 서지정보는 국가과학기술정보센터(NDSL, <http://www.ndsl.kr>)의 논문 고급검색 서비스를 이용하여 엑셀 및 텍스트 파일형태로 저장하고 이를 데이터베이스로 구축하였다. 데이터정제 단계에서 이용한 텍스트마이닝은 자연어처리(Natural language processing) 기술을 이용하여 비정형 혹은 반정형 텍스트 데이터로부터 유용한 정보를 추출·가공하는 것을 목적으로 하는 기술이다. 텍스트마이닝 기술을 통해 방대한 텍스트 문치에서 유의미한 정보를 추출해내고, 다른 정보와의 연계성을 파악하거나 텍스트가 가진 범주를 찾아내는 등 단순한 정보검색 그 이상의 결과를 얻어낼 수 있다. 동시단어분석은 텍스트에 제시된 주제영역내에 존재하는 아이디어들간의 관계를 파악하기 위해 텍스트내에 존재하는 키워드(단어 혹은 명사구)들의 동시출현패턴을 이용하는 내용분석 기법으로 정의할 수 있다. 즉, 논문의 영문초록에 포함된 텍스트를 대상으로 키워드를 생성하고 키워드들간의 동시출현행렬을 생성하여 이를 바탕으로 네트워크 형태로 가시화하는 기법이라고 할 수 있다. 네트워크의 전체 형태와 개별 행위자(키워드)들의 역할을 설명할 수 있는 다양한 지수들을 이용하여 네트워크의 구조와 변화형태를 설명할 수 있다.

키워드 군집분석 단계에서 이용한 군집분석(Cluster analysis)은 어떤 개체나 대상들이 가지고 있는 다양한 특성에 기초하여 동질성을 지닌 군집으로 집단화하는 방법이다. 즉, 군집분석은 분명한 분류기준이 없거나 알려져 있지 않은 상태에서 활용할 수 있는 유용한 분석수단이라고 할 수 있다. 지식맵 분석 단계에서 이용한 지식맵(Knowledge map)은 대량의 연구자원 정보 속에 숨겨진 특별한 형태(Type)와 패턴(Pattern)을 발견하여 그 의미를 파악할 수 있도록 가시적인 형태의 결과로 보여주는 것을 말한다. 이는 대량의 과학정보에 대해 컴퓨터와 계량모형을 활용하여 메타분석을 수행하는 계량정보분석(Informetrics) 결과물의 한 형태로 볼 수 있다. 분석 대상이 되는 연구자원 정보로는 논문, 책, 보고서, 특허, 연구과제 등이 있다. 이러한 자료들의 서지정보를 데이터베이스화하고 텍스트 마이닝, 데이터 마이닝, 계량 지표화 등의 방법으로 계량정보를 분석하게 되면, 자료 전체의 통계적 특성과 이에 기반한 지식 구조를 도출할 수 있으며, 분석결과를 시각화(Visualization)한 것이 바로 지식맵이다(하태정 외, 2011). 지식맵은 분석목적에 따라 크게 3가지 유형으로 분류할 수 있는데, 연구자 네트워크, 저자동시인용 네트워크, 연구주제 네트워크 등이 있다. 결과해석 단계에서는 동시단어분석과 네트워크 지수분석을 통해 가장 파급효과가 큰 연구주제, 연구주제들간의 연관성, 새롭게 부상하거나 쇠퇴하는 연구주제 등의 결과를 도출할 수 있다. 품질경영 분야 연구에서 연구의 주변부에 있던 연구주제들이 중앙부로 이동하는 형태와 미성숙 단계에서 성숙단계로 변화하는 모습을 파악할 수 있으며, 차세대 분야에서 독자적인 분야로 변화하는 연구주제의 역동성을 분석할 수 있다. 또한 2002년 이후 13년 동안 시간의 흐름에 따른 연구주제 영역의 변화형태와 특성을 파악할 수 있다.

4. 연구내용 및 결과

본 연구는 국내 품질경영분야 지식구조 분석을 위해 품질경영학회지에 게재된 논문을 분석대상으로 하였다. 품질경영학회지를 선정한 이유는 품질경영학회지는 1966년 창간호를 시작으로 기업이나 조직의 품질향상 및 품질경영 활동의 효율성 제고를 위해 이론, 기법, 성공사례, 경험 등을 중심으로 학계와 산업계가 활발하게 공동연구를 수행함으로써 경쟁력을 이룰 수 있는 기회를 제공한 국내 품질경영 분야 대표 연구 성과물이기 때문이다. 2002년부터

2014년까지 발행된 논문들 중 영문초록이 존재하는 625개 논문을 선정하고, 서지정보는 국가과학기술정보센터 (NDSL, <http://www.ndsl.kr>)의 논문 고급검색 서비스를 이용하여 엑셀 및 텍스트 파일형태로 저장하고 이를 데이터 베이스로 구축하였다.

텍스트마이닝은 자연어처리(Natural language processing) 기술을 이용하여 비정형 혹은 반정형 텍스트 데이터로부터 유용한 정보를 추출·가공하는 것을 목적으로 하는 기술이다. 텍스트마이닝 기술을 통해 방대한 텍스트 문지에서 유의미한 정보를 추출해내고, 다른 정보와의 연계성을 파악하거나 텍스트가 가진 범주를 찾아내는 등 단순한 정보검색 그 이상의 결과를 얻어낼 수 있다. 본 연구에서는 텍스트마이닝을 위해 R의 tm 패키지를 이용하였다. 읽어 들인 영문초록을 plain text로 전환하고 공백제거, 소문자 변환, 마침표 제거, 불용어 처리, stemming, 동의어 변환 등의 정제를 거친 후 단어들과 문서들간의 관계를 표현하기 위하여 단어문서행렬(Term document matrix)을 생성 하였으며 생성된 행렬은 표 1과 같다.

Table 1. Term document matrix

Document Term	Paper1	Paper2	Paper3	Paper4	Paper5	Paper6	Paper7	Paper8	Paper9	Paper10	...	Paper625
activity	0	0	0	0	0	0	0	0	0	0	...	0
analysis	0	2	0	0	1	3	0	1	0	1	...	1
application	0	0	0	0	0	0	0	0	0	0	...	0
approach	2	3	0	0	0	0	0	0	0	0	...	0
business	0	0	0	0	1	0	0	0	0	2	...	0
capability	0	0	0	0	0	0	0	0	0	0	...	0
case	2	0	0	3	0	0	0	1	3	0	...	1
.
.
way	0	0	0	0	1	0	0	0	0	0	...	0

동시단어분석은 텍스트에 제시된 주제영역내에 존재하는 아이디어들간의 관계를 파악하기 위해 텍스트내에 존재하는 키워드(단어 혹은 명사구)들의 동시출현패턴을 이용하는 내용분석 기법으로 정의할 수 있다. 즉, 논문의 영문초록에 포함된 텍스트를 대상으로 단어를 생성하고 단어들간의 동시출현행렬을 생성하여 이를 바탕으로 네트워크 형태로 가시화하는 기법이라고 할 수 있다. 네트워크의 전체 형태와 개별 행위자(단어)들의 역할을 설명할 수 있는 다양한 지수들을 이용하여 네트워크의 구조와 변화형태를 설명할 수 있다. 단어출현행렬은 논문과 단어간의 출현빈도를 바탕으로 생성된 행렬을 의미하는데 한 단어가 한 논문에 존재하면 1이고 존재하지 않으면 0을 값을 할당하여 행렬을 구성한다. 단어동시출현행렬은 단어출현행렬을 이용하여 행렬계산을 통해 생성할 수 있는데, 출현행렬을 X라고 하고 행렬 X의 전치행렬을 XT라고 하면 동시출현행렬은 X*XT를 통해 구할 수 있다. 단어문서행렬에서 전체 행렬의 원소 수 대비 0의 값을 가지는 원소의 비율을 sparsity로 정의할 수 있는데, 단어문서행렬을 생성하면 905개 단어와 625개 문서로 구성된다. 생성된 행렬의 sparsity가 98%로 매우 높게 나타나 sparsity 89%를 적용하여 106개 단어

로 구성된 단어문서행렬을 생성하였으며 도출된 단어동시출현행렬은 표 2와 같다.

Table 2. Term co-occurrence matrix

Term \ Term	activity	analysis	application	approach	business	capability	case	...	way
activity	57	15	5	8	12	3	8	...	5
analysis	15	168	12	22	15	7	20	...	10
application	5	12	45	6	6	3	10	...	4
approach	8	22	6	77	7	7	10	...	4
business	12	15	6	7	69	3	9	...	10
capability	3	7	3	7	3	32	7	...	1
case	8	20	10	10	9	7	97	...	5
.	
.	
way	5	10	4	4	10	1	5	...	34

단어동시출현행렬로부터 단어별 출현빈도를 정리하면 표 3과 같으며 출현빈도분포를 요약하면 표 4와 같다. 출현 빈도가 가장 높은 단어는 quality와 service로 각각 945회와 627회를 보이고 있다. 그 다음 400회 이상 출현된 단어는 model, process, system, customer로 나타났다. 100회 미만 빈도수를 나타낸 단어의 비율은 27%, 단어수는 60개이고, 100회 이상의 빈도수를 나타낸 단어 비율은 73%, 단어 수는 46개로 나타났다. 출현빈도를 기준으로 단어들을 한눈에 파악할 수 있는 word cloud로 생성하면 그림 2와 같다. Word cloud 중심부에는 출현빈도가 높은 단어들이 위치하고 출현빈도가 낮은 단어들이 중심부로부터 먼 가장 자리에 위치하고 있다. 빈도분석에서 출현빈도가 높게 나타난 quality, service, model, process, system, customer 등의 단어가 중심부에 글씨 크기가 크게 나타나고 있다.

Table 3. Word frequency

No	Word	Frequency	Ratio of frequency	No	Word	Frequency	Ratio of frequency
1	quality	945	6.5	55	need	90	0.6
2	service	627	4.3	56	distribution	88	0.6
3	model	545	3.8	57	measure	88	0.6
4	process	517	3.6	58	procedure	87	0.6
5	system	425	2.9	59	function	86	0.6
6	customer	413	2.8	60	korea	85	0.6
7	factor	373	2.6	61	difference	83	0.6
8	management	368	2.5	62	measurement	79	0.5
9	method	361	2.5	63	mean	79	0.5
10	performance	338	2.3	64	tool	78	0.5
11	company	310	2.1	65	rate	77	0.5
12	data	273	1.9	66	optimal	76	0.5
13	satisfaction	269	1.9	67	capability	74	0.5
14	analysis	255	1.8	68	type	72	0.5

No	Word	Frequency	Ratio of frequency	No	Word	Frequency	Ratio of frequency
15	effect	224	1.5	69	change	71	0.5
16	control	218	1.5	70	sample	70	0.5
17	product	205	1.4	71	production	67	0.5
18	design	204	1.4	72	size	67	0.5
19	level	200	1.4	73	technology	64	0.4
20	sigma	186	1.3	74	environment	63	0.4
21	chart	173	1.2	75	firm	61	0.4
22	industry	168	1.2	76	example	61	0.4
23	cost	156	1.1	77	parameter	61	0.4
24	business	155	1.1	78	requirement	59	0.4
25	variable	154	1.1	79	life	58	0.4
26	activity	146	1.0	80	condition	57	0.4
27	test	145	1.0	81	goods	56	0.4
28	six	144	1.0	82	methodology	56	0.4
29	relationship	138	0.9	83	component	56	0.4
30	standard	133	0.9	84	number	55	0.4
31	time	129	0.9	85	concept	54	0.4
32	value	129	0.9	86	market	54	0.4
33	manufacturing	127	0.9	87	application	54	0.4
34	reliability	126	0.9	88	point	52	0.4
35	korean	122	0.8	89	equation	51	0.4
36	case	120	0.8	90	importance	51	0.4
37	order	118	0.8	91	field	51	0.4
38	improvement	114	0.8	92	simulation	50	0.3
39	strategy	113	0.8	93	issue	48	0.3
40	problem	112	0.8	94	decision	47	0.3
41	approach	110	0.8	95	impact	47	0.3
42	information	108	0.7	96	implementation	47	0.3
43	survey	105	0.7	97	plan	45	0.3
44	efficiency	104	0.7	98	set	45	0.3
45	use	101	0.7	99	technique	45	0.3
46	project	100	0.7	100	term	45	0.3
47	evaluation	99	0.7	101	program	44	0.3
48	iso	99	0.7	102	regression	44	0.3
49	index	98	0.7	103	operation	43	0.3
50	failure	96	0.7	104	questionnaire	40	0.3
51	organization	96	0.7	105	key	37	0.3
52	development	94	0.6	106	way	37	0.3
53	characteristic	93	0.6				
54	influence	92	0.6				

Table 4. Summary of word frequency

Frequency	< 100	≥ 100	Sum	≥ 200	≥ 500	Sum
Frequency ratio(%)	27.0	73.0	100.0	48.7	18.1	66.8
Number of words	60	46	106	19	4	23
Ratio of words(%)	56.6	43.4	100.0	17.9	3.8	21.7



Figure 2. Word cloud

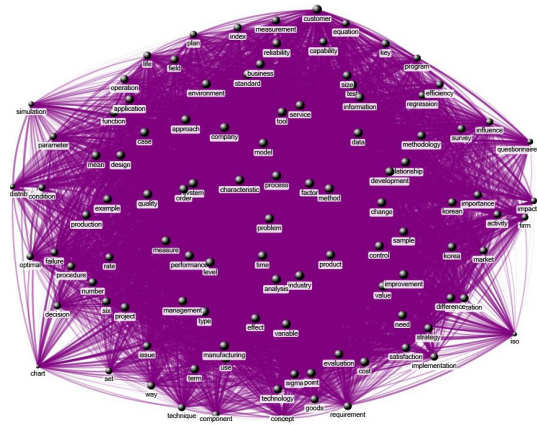


Figure 3. Word network diagram

추출된 106개 단어를 대상으로 NodeXL을 이용하여 네트워크 분석을 수행한 결과는 그림 3과 같다. 생성된 네트워크는 106개의 노드와 5,314개의 링크를 가지며 네트워크 밀도는 0.95로 분석되었다. 또한 노드간의 평균거리는 1.04이고 2개의 노드만 거치면 모두 연계가 가능한 네트워크로 나타났다. 평균 매개 중심성지수는 2.37, 평균 근접 중심성지수는 0.01, 평균 위세 중심성 지수는 0.01로 분석되었다. 네트워크의 가독성을 높이기 위하여 연결중심성 지수가 40보다 큰 링크만 네트워크상에 표시하면 전체 단어네트워크는 그림 4와 같이 형성된다.

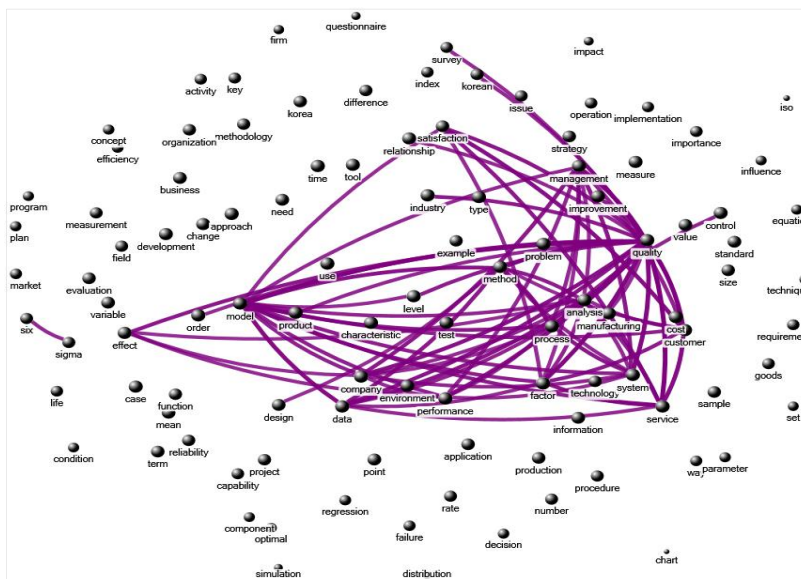


Figure 4. Word network diagram(degree>40)

생성된 단어 네트워크의 연결 중심성 지수를 분석해보면 최대 연결 중심성은 105이고 최소 연결 중심성은 80으로 나타났다. 최대 연결 중심성 지수 105를 가지는 24개 단어와 단어의 빈도수, 최대 연관성을 가지는 단어와 최대 상관계수를 정리하면 다음과 같다. 단어들간의 상관관계(correlation)는 단어동시출현행렬(C_{ij})로부터 계산할 수 있다. 단어 i 의 출현빈도를 C_{ii} , 단어 j 의 발생빈도를 C_{jj} , 단어 i 와 j 의 동시발생빈도를 C_{ij} 라고 할 때, 단어 i 와 j 의 상관관계(correlation)는 단어 i 와 j 가 동시에 출현할 확률 P_{ij} 로 주어진다.

$$P_{ij} = \frac{C_{ij}}{C_{ii} + C_{jj} - C_{ij}}$$

연결중심성(degree centrality) 지수가 가장 높은 105로 나타난 24개 단어와 빈도수, 이들 단어와 상관관계가 가장 높은 단어와 상관계수를 정리하면 표 5와 같다. 빈도분석에서 빈도가 높게 나타났던 quality, service, model, process, system, customer 단어의 연결중심성도 105로 가장 높게 나타났다. 연결중심성 지수가 높다는 것은 다양한 다른 단어들과 많은 연관성을 가지고 있다는 것을 의미한다. quality와 가장 연관성이 높은 단어는 service이고, model과 가장 연관성이 높은 단어는 variable로 나타났다. Process와 가장 연관성이 높은 단어는 control이고, system과 가장 연관성이 높은 단어는 iso이며, customer와 가장 연관성이 높은 단어는 satisfaction으로 분석되었다.

Table 5. Degree centrality & correlation

Word	Degree	Frequency	Max correlation word	Max correlation
quality	105	945	service	0.38
service	105	627	quality	0.38
model	105	545	variable	0.31
process	105	517	control	0.27
system	105	425	iso	0.26
customer	105	413	satisfaction	0.58
factor	105	373	satisfaction	0.24
method	105	361	design	0.16
performance	105	338	effect	0.2
company	105	310	management	0.34
data	105	273	analysis	0.2
analysis	105	255	regression	0.22
level	105	200	service	0.28
industry	105	168	issue	0.28
cost	105	156	failure	0.21
time	105	129	failure	0.24
manufacturing	105	127	company	0.16

Word	Degree	Frequency	Max correlation word	Max correlation
order	105	118	efficiency	0.11
problem	105	112	technique	0.23
approach	105	110	methodology	0.2
use	105	101	level	0.21
characteristic	105	93	concept	0.15
type	105	72	key	0.15
product	105	67	design	0.17

네트워크 분석을 통해 전체 네트워크의 구조와 특성을 파악할 수 있지만, 일반적으로 단어 수가 수십 개에서 수백 개 도출되므로 전체를 동시에 해석하는 데는 무리가 따를 수 있다. 따라서 단어들을 단어들간의 유사성이나 거리를 이용하여 비슷한 특성을 지닌 몇 개 군집으로 집단화할 필요가 있다. 군집분석(Cluster analysis)은 어떤 개체나 대상들이 가지고 있는 다양한 특성에 기초하여 동질성을 지닌 군집으로 집단화하는 방법이다. 즉, 군집분석은 분명한 분류기준이 없거나 알려져 있지 않은 상태에서 활용할 수 있는 유용한 분석수단이라고 할 수 있다.

본 연구에서 군집분석 시 활용한 K-means 알고리즘은 입력 값으로 군집수(k)를 취하면서 군집내 유사성은 높고 군집들 사이의 유사성은 낮도록 n개의 객체 집합을 k개의 군집으로 분류하는 방법이다. 군집유사성은 군집에서 군집의 평균값을 기준으로 산정하는데 이는 군집의 무게중심으로 볼 수 있기 때문이다. K-means 클러스터링 알고리즘 적용 시 문제가 되는 점은 군집 수를 사전에 입력 값으로 제공해야 한다는 점인데, 본 연구에서는 최적 군집 수 결정을 위해 Calinski-Harabasz index와 average silhouette width index를 이용하였으며, 두 개 지수를 간략하게 설명하면 다음과 같다.

Callinski-Harabasz index(Zumel & Mount, 2014)는 군집간분산과 군집내분산의 비율을 의미한다. 군집간분산 B는 $BSS(k)/(k-1)$ 로 주어지고, 군집내분산 W는 $WSS(k)/(n-k)$ 로 주어진다. $WSS(k)$ 는 하나의 군집에 속하는 데이터들이 얼마나 근접한지를 의미하고, $BSS(k)$ 는 군집들이 얼마나 멀리 떨어져 있는가를 의미한다. $WSS(k)$ 가 작고 $BSS(k)$ 는 클수록 좋은 클러스터링이라고 할 수 있으며 Callinski-Harabasz index(CH)는 다음과 같이 주어진다.

$$CH = \frac{B}{W} = \frac{\frac{BSS(k)}{k-1}}{\frac{WSS(k)}{n-k}} = \frac{n-k}{k-1} \frac{BSS(k)}{WSS(k)}$$

Silhouette width는 군집의 강도를 측정하는 방법으로 개별 원소(element)가 얼마나 잘 군집화 되었는가를 의미한다(Kaufman & Rousseeuw 1990). a_i 를 데이터 집합을 구성하는 점 X_i 와 i 번째 군집에 속하는 모든 다른 점들간의 거리평균이라고 하고, b_i 를 점 X_i 와 다른 군집에 속하는 모든 점들까지 최소 평균거리라고 하면 silhouette width

$$SW_i = \frac{b_i - a_i}{\max(a_i, b_i)}$$

로 주어지고, average silhouette width는 $\frac{\sum_i SW_i}{n}$ 으로 주어진다.

K-means 군집분석 알고리즘을 이용하여 군집의 수를 1부터 10까지 변화시키면서 Calinski-Harabasz index(ch)와 average silhouette width index(asw)를 그래프로 나타내면 그림 5와 같다. 그림을 살펴보면 최적 군집

수는 7개라는 것을 확인할 수 있다.

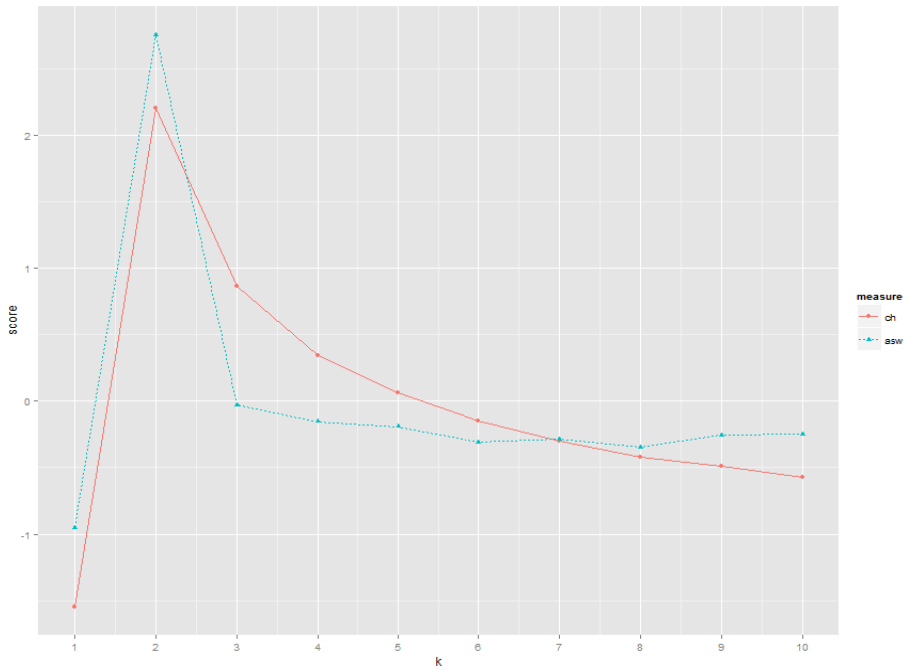


Figure 5. Calinski-Harabasz and average silhouette width indices

최적 군집수로 결정된 7개 군집을 입력 값으로 하여 군집분석을 수행하면 표 6과 같으며, 군집별로 빈도수가 높은 상위 3개 단어와 이들 단어와 연관성이 가장 높은 단어를 추출하고 전문가 협의를 통해 군집명을 부여하였다. 군집1에 속하는 빈도수 상위 3개 단어는 standard, efficiency, evaluation이고 연관성이 가장 높은 단어들을 고려해 볼 때 iso와 같은 표준적용과 평가모형을 통한 효율성 측정과 관련된 연구주제들이 연관된 것으로 판단되어 standard & efficiency로 군집명을 부여하였다. 군집2는 product, design, level이란 단어가 빈도수가 높은 단어로 나타났으며 제품기능설계나 서비스수준 향상을 위한 연구주제와 연관성이 높은 것으로 판단하여 product design으로 군집명을 부여하였다. 군집3의 고빈도어는 reliability, failure, distribution인데 이들 단어들은 전통적으로 신뢰성과 관련이 높은 연구주제이므로 군집명을 reliability로 부여하였다. 군집4의 고빈도어는 control, chart, variable로 품질경영 분야의 전통적인 연구주제인 관리도와 연관되므로 군집명을 control chart로 부여하였다. 군집5의 고빈도어는 quality, model, process로 나타났으며 품질모형과 공정관리와 관련이 있는 연구주제로 판단되어 quality model로 군집명을 부여하였다. 군집6의 고빈도어는 satisfaction, sigma, industry로 6sigma를 통한 고객만족이란 연구주제와 연관성이 있는 것으로 판단하여 6sigma로 군집명을 부여하였다. 마지막으로 군집7의 고빈도어는 service, system, customer로 고객만족을 통한 서비스품질과 연관성이 있는 것으로 판단하여 군집명으로 service quality를 부여하였다.

Table 6. Results of cluster analysis

Cluster	Word	Frequency	Max correlation word	Cluster name
Cluster1	standard	133	iso	Standard & efficiency
	efficiency	104	measure	

Cluster	Word	Frequency	Max correlation word	Cluster name
	evaluation	99	model	
Cluster2	product	205	design	Product design
	design	204	function	
	level	200	service	
Cluster3	reliability	126	field	Reliability
	failure	96	time	
	distribution	88	simulation	
Cluster4	control	218	chart	Control chart
	chart	173	control	
	variable	154	model	
Cluster5	quality	945	service	Quality model
	model	545	variable	
	process	517	control	
Cluster6	satisfaction	269	customer	6sigma
	sigma	186	six	
	industry	168	issue	
Cluster7	service	627	quality	Service quality
	system	425	iso	
	customer	413	satisfaction	

전략맵(Strategic diagram)은 단어동시출현행렬로부터 개별 단어의 centrality와 density를 계산하여 2차원상에 단어의 분포나 단어군집을 표시한 것이다. Centrality는 개별 키워드가 군집 외부의 다른 단어들과 얼마나 강한 연결 관계를 갖는지를 계산한 값이고, density는 개별 단어가 동일 군집에 속하는 다른 단어들과 얼마나 강한 연결 관계를 갖는지를 나타낸다. 전략맵을 생성하기 위해서는 단어동시출현행렬로부터 두 단어간의 연결의 강도를 나타내는 equivalence index(E_{ij})를 계산해야 하는데 다음과 같이 주어진다.

$$E_{ij} = \frac{c_{ij}^2}{c_i \cdot c_j}$$

여기서, c_{ij} 는 단어 i 와 단어 j 가 동시에 출현하는 문서의 수를 의미하고, c_i 와 c_j 는 개별 단어가 출현하는 문서의 수를 의미한다(Callon et al. 1991). 표준화된 단어동시출현행렬로부터 centrality와 density를 계산해야 하는데 계산방법은 연구자마다 다양하게 정의하여 이용되고 있다. 본 연구에서는 정용일 외(2005)가 제안한 centrality와 density 계산방법을 이용하였는데 다음과 같이 주어진다.

$$Centrality = \frac{\sum E_{ij}}{N-n}, \quad Density = \frac{\sum E_{ij}}{n-1}$$

여기서, E_{ij} 는 equivalence index, n 은 동일한 군집에 속하는 단어 수, N 은 전체 단어 수를 의미한다. Centrality는 군집 외부의 E_{ij} 값의 합을 자유도 $N-n$ 으로 나눈 값으로 계산되고, density는 군집 내부의 E_{ij} 값의 합을 자유도 (최대 연결가능한 단어 수) $n-1$ 로 나눈 값으로 계산된다. 본 연구에서는 2차원상에 표현할 군집의 크기(원이나 구의 크기)는 전체 단어가 문서(논문)에서 출현한 빈도수 대비 군집내에 속하는 단어가 출현한 문서 수의 상대적인 비율(RF_i)로 계산하였다.

$$RF_i = \frac{\sum_{j \in i} F_j}{\sum_{j \in \forall i} F_j}, \quad i=1, 2, \dots, k, \quad j=1, 2, \dots, N$$

여기서, $\sum_{j \in i} F_j$ 는 i 군집에 속하는 단어가 출현하는 문서 수의 합을 나타내고, i 는 군집의 수, j 는 단어를 의미한다.

전략맵의 해석은 그림 6에서 보는 바와 같이 1사분면에 있는 군집은 중심성과 밀도가 모두 높아 가장 주류를 이루고 있는 연구분야를 의미하고, 2사분면에 위치한 군집은 내부 성숙도는 높으나 중심성이 낮아 자체적으로 독자적 완성도를 갖추면서 주목을 덜 받게 되는 분야를 의미한다. 3사분면에 위치하는 군집은 주변부이고 미성숙한 분야를 의미하고, 밀도는 낮으나 중심성이 높은 4사분면에 있는 군집은 주목을 끄는 분야로 활발히 논의되고 있는 중심적인 위치를 차지하고는 있으나 아직 내부적 성숙도는 부족한 분야를 나타낸다(Jung et al. 2005).

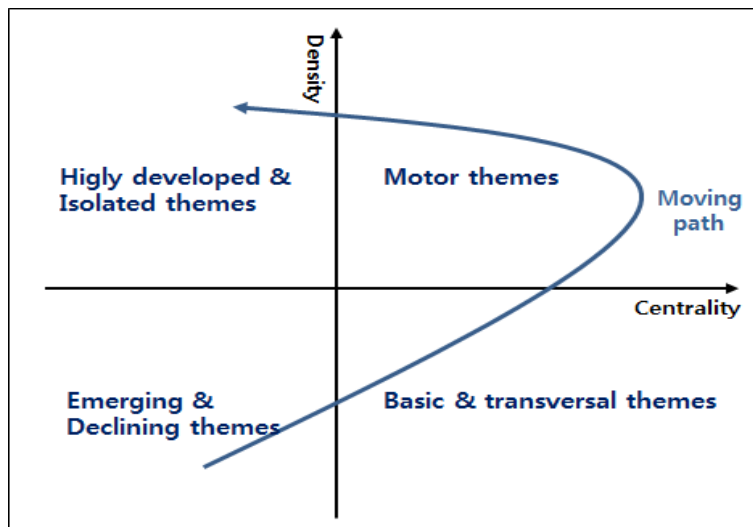


Figure 6. Strategic diagram(Lopez-Herrera et al.)

전체 106개 단어를 대상으로 계산된 centrality와 density를 기준으로 전체 단어를 전략맵상에 표시하면 그림 7

과 같다. 그림 7은 단어들의 전체 분포를 개괄적으로 파악할 수 있는 있지만 단어가 속한 군집의 위치와 군집의 상대적 비중을 파악하기에는 무리가 있다. 따라서 군집분석을 통해 도출된 군집을 기준으로 전략맵을 생성할 필요가 있다. 군집별 centrality와 density를 계산하고 해당 군집에 속하는 단어수와 단어가 출현한 논문수, 논문수의 상대적 비중, 상대적 비중순위, 사분면 값을 계산하고 정리하면 표 7과 같다.

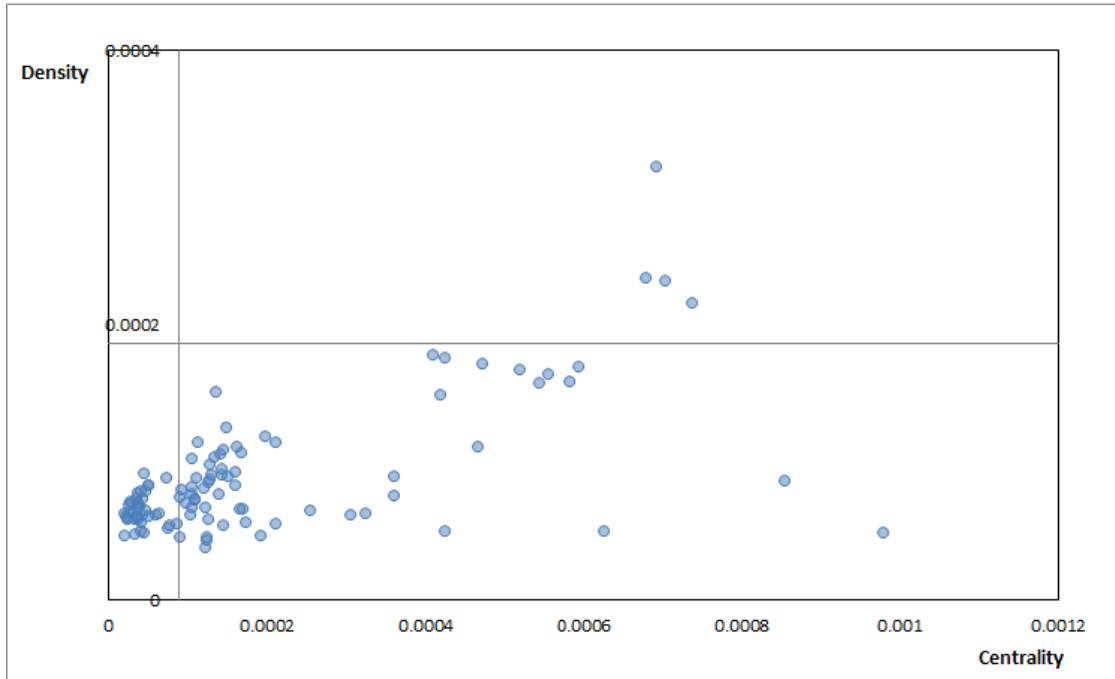


Figure 7. Strategic diagram of all words

Table 7. Centrality & density by cluster(2002~2014)

Cluster name	Centrality	Density	Number of papers	Number of words	Ratio of paper frequency	Rank	Quadrant
Standard & efficiency	0.000067	0.000038	1,426	36	0.192	2	III
Product design	0.000102	0.00015	1,121	13	0.151	4	IV
Reliability	0.000057	0.000135	732	16	0.098	6	IV
Control chart	0.000068	0.000599	345	6	0.046	7	IV
Quality model	0.00025	0.000702	905	4	0.122	5	I
6sigma	0.000085	0.000141	1,393	21	0.187	3	IV
Service quality	0.000161	0.000497	1,524	10	0.205	1	IV
Average/Sum	0.000187	0.000089	7,446	106	1.000	-	-

시간의 변화에 따른 연구주제의 변화형태를 살펴보기 위하여 2002년부터 2014년까지 발표된 논문의 전체 분석

기간을 2002년부터 2007년까지 게재된 300편의 논문을 전반기로 2008년부터 2014년까지 게재된 325개 논문을 후반기로 구분하여 전략맵 분석을 수행하였다. 전반기와 후반기의 군집별 분석결과는 표 8과 표 9와 같다. 전반기와 후반기 데이터의 경우 최적 군집 수는 모두 8개로 결정되었으며, 전반기 군집은 control chart, standard & internet business, service quality, product design, customer satisfaction, process model, reliability, 6sigma로 군집화되었다. 후반기 데이터의 경우 system performance, reliability, 6sigma, service quality, product design, management innovation, customer satisfaction, control chart 등으로 군집화되었다. 표 7, 표 8, 표 9의 분석결과를 바탕으로 전략맵을 생성하면 그림 8, 그림 9, 그림 10과 같다. 그림 8은 2002년부터 2007년까지의 데이터를 이용하여 전략맵으로 표현한 결과이고, 그림 9는 2008년부터 2014년까지의 데이터를 기반으로 전략맵을 생성한 결과이다. 또한 그림 10은 2002년부터 2014년까지 전체 분석대상 데이터를 기준으로 전략맵을 분석한 결과이다.

Table 8. Centrality & density by cluster(2002~2007)

Cluster name	Centrality	Density	Number of papers	Number of words	Ratio of paper frequency	Rank	Quadrant
Control Chart	0.000534	0.000063	301	12	0.0840	6	IV
Standard & Internet business	0.000052	0.000075	649	32	0.1811	1	III
Service quality	0.001139	0.000234	380	4	0.1061	5	IV
Product design	0.000178	0.000105	471	12	0.1315	4	IV
Customer satisfaction	0.000467	0.000139	611	10	0.1705	2	IV
Process model	0.000628	0.000236	292	3	0.0815	7	IV
Reliability	0.000263	0.00007	284	11	0.0793	8	IV
6sigma	0.000164	0.000082	595	21	0.1661	3	IV
Average/Sum	0.000095	0.000263	3,583	105	1.0000	-	-

Table 9. Centrality & density by cluster(2008~2014)

Cluster name	Centrality	Density	Number of papers	Number of words	Ratio of paper frequency	Rank	Quadrant
System performance	0.00055	0.000165	295	4	0.0743	8	IV
Reliability	0.000145	0.000057	468	19	0.1179	4	III
6sigma	0.000149	0.000063	434	19	0.1094	6	III
Service quality	0.000704	0.000227	683	6	0.1721	1	I
Product design	0.000153	0.000108	643	14	0.1621	2	IV
Management innovation	0.000099	0.000081	581	18	0.1464	3	III
Customer satisfaction	0.00082	0.000144	421	6	0.1061	7	IV
Control chart	0.000073	0.000062	443	21	0.1116	5	III

Average/Sum	0.000088	0.000209	3,968	107	1.0000	-	-
-------------	----------	----------	-------	-----	--------	---	---

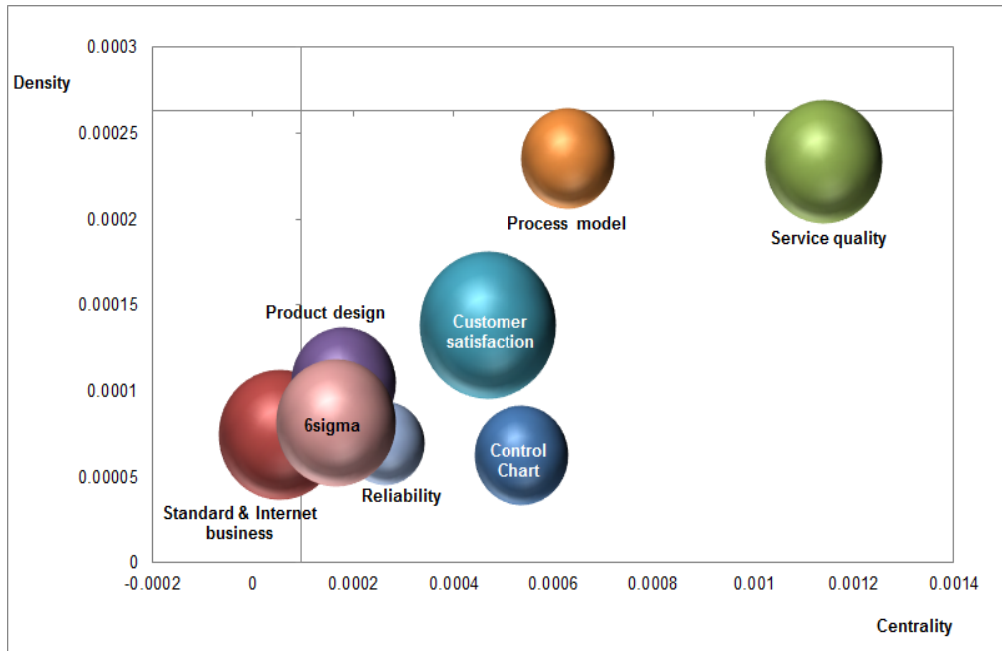


Figure 8. Strategic diagram(2002~2007)

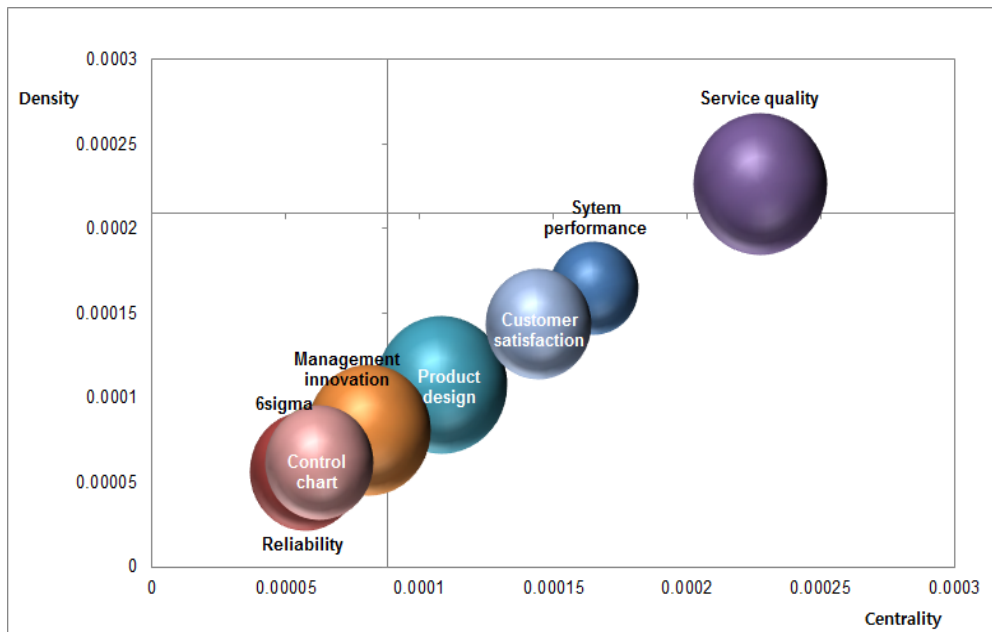


Figure 9. Strategic diagram(2008~2014)

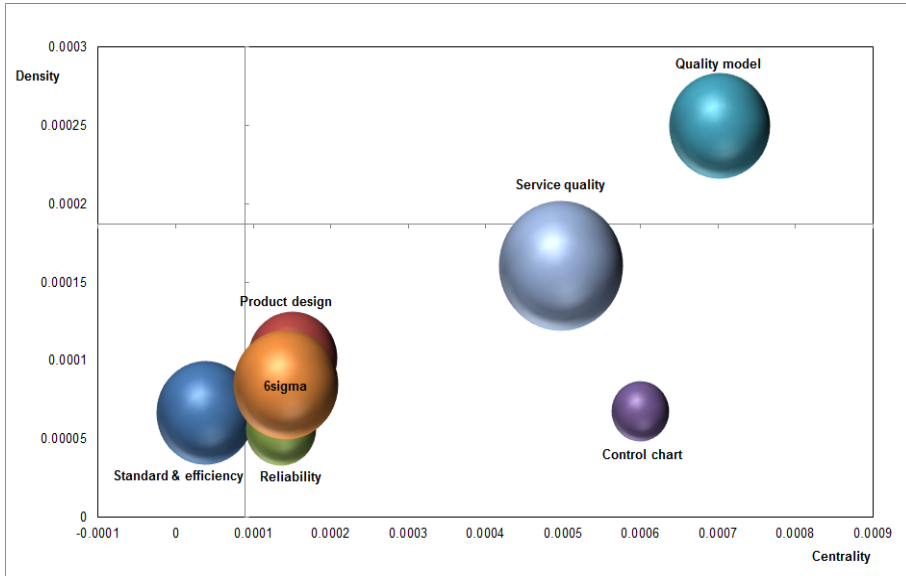


Figure 10. Strategic diagram(2002~2014)

전체 기간을 대상으로 한 전략맵 분석결과에 따르면 표준화와 효율성(Standard & efficiency)이 새롭게 부상한 연구주제로 나타났으며, 6시그마(6sigma), 신뢰성(Reliability), 관리도(Control chart) 관련 연구주제는 상대적으로 연구비중이 감소하고 있다. 제품설계(Product design)와 서비스품질(Service quality) 관련 연구주제의 상대적인 중요성이 높아지고 있으며, 품질모델(Quality model) 관련 연구주제가 핵심적인 연구주제로 분석되었다.

전반기와 후반기 전략맵을 비교분석해보면 품질경영분야의 전통적인 연구주제인 신뢰성(Reliability), 6시그마(6sigma), 관리도(Control chart) 관련 연구주제는 3사분면에 속하고 비중이 감소하는 것으로 볼 때 관련 연구의 중요성이 쇠퇴하고 있는 것으로 분석된다. 제품설계(Product design)와 고객만족(Customer satisfaction) 관련 연구주제는 전반기와 동일하게 중요한 연구주제로 자리잡고 있다. 경영혁신(Management innovation) 관련 연구주제는 새롭게 부상하고 있으며, 프로세스 모델(Process model) 관련 연구주제는 시스템 성과(System performance) 관련 연구주제로 연구범위가 확대된 것으로 분석된다. 1사분면에 위치한 서비스 품질(Service quality) 관련 연구주제가 핵심 연구주제로 부상하고 있는 것으로 나타났으며, 이는 제조업 중심에서 서비스업의 중요성이 부각되는 산업구조 변화와 관련성이 있는 것으로 판단된다.

이상의 분석결과를 종합해볼 때 서비스품질(Service quality) 관련 연구주제의 중요도와 비중이 증가할 것으로 판단되고, 제품설계(Product design)와 시스템성과(System performance) 관련 연구도 증가할 것으로 예상된다. 산업구조가 제조업 중심에서 서비스업 중심으로 변화하고 있고, 시스템의 경우 부분 최적화에서 전체 최적화를 추구하는 방향으로 변화하고 있는 산업의 패러다임 변화가 반영된 결과라고 분석된다.

5. 결 론

본 연구에서는 품질경영분야 지식구조의 변화행태와 연구주제의 변화추세 분석을 위하여 동시단어분석과 군집분

석, 전략맵 등의 기법을 이용하여 단어들의 네트워크 구조와 지식구조를 맵 형태로 가시화하였다. 본 연구를 통해 도출된 연구결과를 정리하면 다음과 같다.

첫째, 단어동시출현행렬로부터 생성된 단어네트워크는 106개의 노드와 5,314개의 링크를 가지며 네트워크 밀도는 0.95로 분석되었다. 네트워크의 평균 매개 중심성지수는 2.37, 평균 근접 중심성지수와 평균 위세 중심성 지수는 0.01로 분석되었다. 둘째, 단어동시출현행렬로부터 최적 군집수 결정기준과 K-means 군집분석 적용을 통해 106개 단어가 standard & efficiency, product design, reliability, control chart, quality model, 6sigma, service quality 등 7개 군집으로 형성되었다. 셋째, 시간의 흐름에 따른 전략맵 분석결과에 따르면 품질경영분야의 전통적인 연구 주제인 신뢰성(Reliability), 6시그마(6sigma), 관리도(Control chart) 관련 연구주제는 3사분면에 속하고 비중이 감소하는 것으로 볼 때 관련 연구의 중요성이 쇠퇴하고 있는 것으로 분석된다. 제품설계(Product design)와 고객만족(Customer satisfaction) 관련 연구주제는 전체 분석기간에 걸쳐 중요한 연구주제로 자리잡고 있다. 경영혁신(Management innovation) 관련 연구주제는 새롭게 부상하고 있으며, 프로세스 모델(Process model) 관련 연구주제는 시스템 성과(System performance) 관련 연구주제로 연구범위가 확대된 것으로 분석된다. 1사분면에 위치한 서비스 품질(Service quality) 관련 연구주제는 핵심 연구주제로 부상하고 있는 것으로 나타났다.

본 연구는 동시단어분석을 이용하여 품질경영 분야 지식구조 변화분석을 수행한 최초의 연구이며, 본 연구결과를 바탕으로 해외 품질경영 분야 지식구조 변화에 대한 비교연구가 가능하고, 타 분야의 지식구조 변화분석에도 활용이 가능할 것으로 판단된다. 본 연구결과와 해외저널에 게재된 품질경영분야 지식구조 분석결과와의 비교분석을 통해 연구분야와 연구주제에 대한 벤치마킹이 가능하고, 이를 바탕으로 연구자에게 미래 유망 연구주제를 제공하고 연구주제 선정을 위한 가이드라인을 제시할 수 있을 것으로 기대된다.

REFERENCES

- ASQ. 2013. The ASQ Global State of Quality Research. Discoveries, 1-39.
- BIA, Future Trends in Quality Management, <http://www.bia.ca>.
- Brass, D. J. 2003. Networks in the Knowledge Economy. Oxford University Press.
- Calinski, T., and Harabasz, J. 1974. "A dendrite method for cluster analysis." Communications in Statistics 3(1):1-27.
- Callon, M., Courtial, J. P., and Laville, F. 1991. "Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry." Scienlomtrirs 22(1):155-205.
- 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 Fuzzy Sets Theory field." Journal of Informatics 5:146-166.
- Feinerer, I. 2008. A text mining framework in R and its applications. Doctoral thesis, WU Vienna University of Economics and Business.
- Feinerer, I. 2008. An Introduction to Text Mining in R. R News 8(2):19-22.
- Feinerer, I., Hornik, K., and Meyer, D. 2008. "Text Mining Infrastructure in R." Journal of statistical software 25(5):1-48,
- Ha, T. J., Lee, J. O., Park, B. W., Suh, J. Y. 2011. An Industrial Map of New Growth Engine: Focused on LED

- and Photovoltaic Industry. STEPI.
- He, Q. 1999. "Knowledge discovery through co-word analysis." *Library Trends* 48(1):133–159.
- JoongAng Ilbo 2011. two-fold increase every two years, Worldwide data Big Bang. Last modified June 30, <http://news.joins.com/article/5713318?cloc=joongang%7Carticle%7Cissue>.
- Jung, Y. I., Lee, J. Y., Lee, B. R., Yoo, S. H., Won, D. K., Jung., S. C., and Joo, S. H. 2005. Mapping and utilization of knowledge through informetric data analysis. KISTI.
- Kaufman, L., Rousseeuw, P. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley.
- Kim, C. M., Jung, U., Seo, J. H., and Bae, S. M. 2016. "Literature Review on the Quality Innovation in KSQM for 50 Years." *Journal of the Korean Society for Quality Management* 44(1):17–28.
- Kim, H. J., and Song, M. 2014. "A Study on the Research Trends in Domestic/International Information Science Articles by Co-word Analysis." *Journal of the Korean Society for Information Management* 31(1):99–118.
- Kwon, H. M., Hong, S. H., Lee, M. K., and Lim, S. U. 2016. "Literature Review on the Statistical Quality Control in Journal of the KSQM for 50 Years." *Journal of the Korean Society for Quality Management* 44(1):1–16.
- Law, J., and Whittaker, J. 1992. "Mapping acidification research: A test of the co-word method." *Scientometrics* 23(3):417–461.
- Lee, D. H., Kim, Y. S., Lee, D. W., and Cho, J. R. 2012. "The Evolution of Research on Quality: Papers Published in Journal of the Korean Society for Quality Management (2007–2011)." *Journal of the Korean Society for Quality Management* 40(2):94–105.
- Lee, J. H. 2011. "A Study on Knowledge Structure of Budget Theory Research Program: Focusing on the paper content analysis and network analysis." *Korean Journal of Policy Analysis and Evaluation* 21(2):115–148.
- Lee, S. K., J, S., Kim, H. G., and Yom, Y. H. 2011. "A Social Network Analysis of Research Topics in Korean Nursing Science." *Journal of Korean Academy Nursing* 41(5):623–632.
- Lee, W. H., Kim, H. J., and Park, J. C. 2008. "Detecting Emerging Technology using Information Analysis." *The Journal of Information Systems* 17(3):235–254.
- Lopez-Herrera, A. G., Cobo, M. J., Herrera-Viedma, E., Herrera, F., Bailon-Moreno, R., and Jimenez-Contreras, E. 2009. "Visualization and evolution of the scientific structure of fuzzy sets research in Spain." *Information research* 14(4):1–23.
- NodeXL, <http://nodexl.codeplex.com>.
- Song, I. S., Park, M. H., Lee, B. R., and Won, D. K. 2006. *Theory of Social Network Analysis and case studies*. KISTI.
- The R Project for Statistical Computing, <http://www.r-project.org>.
- The wall street journal. 2014. Toyota in \$1.1 Billion Gas-Pedal Settlement.
- Yoshimoto, K. 2010. *The Impact and Challenges of Toyota Recall*. Korea-Japan Cooperation Foundation for Industry and Technology.
- Zumel, N., and Mount, J. 2014. *Practical Data Science with R*, Manning Publications.