



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

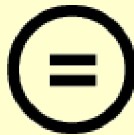
다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학석사 학위논문

토픽모델링과 연관규칙을 활용한 보유 기술
기반의 잠재적 기술테마 탐색



2017년 2월

부경대학교 대학원

시스템경영공학과

김 주 성

공학석사 학위논문

토픽모델링과 연관규칙을 활용한 보유 기술 기반의 잠재적 기술테마 탐색

지도교수 서 원 철

이 논문을 공학석사 학위논문으로 제출함.

2017년 2월

부경대학교 대학원

시스템경영공학과

김 주 성

김주성의 공학석사 학위논문을 인준함.

2017년 2월 24일

위 원 장 공학박사 구 평 회 (인)

위 원 공학박사 서 원 철 (인)

위 원 공학박사 신 승 준 (인)

Identifying potential technology themes based on internal capabilities
using topic modeling and association rule mining

Ju Sung Kim

Division of Systems Management and Engineering, The Graduate School,
The GPukyong National University

Abstract

In a rapidly changing market environment, it is very important to discover new technology opportunities due to global technology competition. Identifying technology opportunities can be defined as the process of discovering opportunities for profitability through the development and utilization of technology and products, and these activities have become essential elements for the continued growth of enterprises and countries. For this reason, many researchers have proposed various methodologies for discovering technology opportunities, but most studies fail to identify the internal competencies of those seeking technology opportunities, or proposed technology opportunities as set items, such as patent classification codes. Therefore, it has been difficult for research results to be used in decision making by corporations or public agencies. In this study, we propose a methodology for recommending non - retention technologies that are highly related to retention technologies, considering the internal competencies of those who seek out technology opportunities using topic modeling and association rule mining. The results of this study are expected to save a lot of time and money in mergers and acquisitions, expansion of business scope, and R & D and policy decision making by government and public research institutes.

목 차

제 1장 서론 -----	01
1.1 연구 배경 -----	01
1.2 연구 목표 -----	02
1.3 연구 프로세스 -----	03
제 2장 배경연구 -----	05
2.1 기술기회 발굴 -----	05
2.2 LDA(Latent Dirichlet Allocation) -----	08
2.3 ARM(Association Rule Mining) -----	11
제 3장 연구 방법론 -----	14
3.1 내부역량 기반의 기술테마 탐색 -----	14
3.1.1 데이터 수집 및 내부기술테마 도출 -----	14
3.1.2 외부기술테마 도출 -----	15
3.2 내부기술테마와 외부기술테마 간의 유사도 산출 --	17
3.3 기술테마 간의 연관규칙 도출 -----	18
3.3.1 문서-기술테마 매트릭스 -----	18
3.3.2 기술테마 간의 연관규칙 도출 -----	19
3.4 잠재적 기술테마 추천 -----	19
3.4.1 최종 추천 네트워크 -----	19
3.4.2 결과 해석 -----	21
3.4.3 상세 특허 추천 -----	21
제 4장 사례연구 -----	22

4.1 내부역량 기반의 기술테마 탐색 -----	22
4.1.1 데이터 수집 및 내부기술테마 도출 -----	22
4.1.2 외부기술테마 도출 -----	24
4.2 내부기술테마와 외부기술테마 간의 유사도 산출 --	26
4.3 기술테마 간의 연관규칙 도출 -----	28
4.3.1 문서-기술테마 매트릭스 -----	28
4.3.2 기술테마 간의 연관규칙 도출 -----	30
4.4 잠재적 기술테마 추천 -----	31
4.4.1 최종 추천 네트워크 -----	31
4.4.2 결과 해석 -----	33
4.4.3 상세 특허 추천 -----	38
제 5장 결론 -----	40
참고 문헌 -----	43
부 록 -----	49
부록.1 내부기술테마 -----	49
부록.2 외부기술테마 -----	55
부록.3 ARM 결과 -----	61
부록.4 LDA R코드 -----	64
부록.5 ARM R코드 -----	67
부록.6 유사도 계산 python코드 -----	68

그림 목차

[그림 2.1] Latent Dirichlet Allocation -----	08
[그림 2.1] LDA 도식화 -----	11
[그림 3.1] 연구프로세스 -----	14
[그림 3.2] 추천 네트워크 구성 -----	20
[그림 4.1] 외부기술 탐색을 위한 쿼리문 -----	24
[그림 4.2] 유사도 산출 테이블 -----	26
[그림 4.3] 문서-토픽 가중치 매트릭스(일부) -----	28
[그림 4.4] 문서-기술테마 매트릭스(일부) -----	29
[그림 4.5] 최종 추천 네트워크 -----	32

표 목차

[표 2.1] IPC 특허분류 코드	07
[표 2.2] ARM 지표	12
[표 3.1] 외부기술 탐색을 위한 쿼리문 작성 예시	16
[표 4.1] 내부기술테마(일부)	23
[표 4.2] 외부기술테마(일부)	25
[표 4.3] 유사도 산출 결과	27
[표 4.4] ARM 결과(일부)	30
[표 4.5] 내부기술테마와 외부기술테마의 단어 비교	33
[표 4.6] 내부기술테마 명칭	34
[표 4.7] 외부기술테마 명칭	35
[표 4.8] 추천 특허 리스트	38

1.서론

1.1 연구배경

최근 글로벌 기술경쟁으로 인해 급변하는 시장 환경 속에서 하나의 기업이 글로벌 마켓을 선점하고 지속 적으로 경쟁 우위를 점하는 것은 매우 어려워 졌다 (노대민 외, 2013). 이에 따라 새로운 기술기회를 탐색 하고 발굴 하여 새로운 시장을 포착하는 활동은 기업 및 국가의 기술 경쟁력을 지속적으로 성장시키기 위한 필수 요소가 되었다(Klevorick et al, 1995), (Abernathy et al, 1985), (Christensen et al, 2013).

기술기회 발굴 이란 ‘기술 및 제품의 개발과 활용을 통해 수익 창출이 가능한 기회를 발굴하는 과정’으로 정의 되어질 수 있으며(M.Lee et al, 2013), 급변 하는 시장 환경 에서 새로운 기술을 발굴하고 선점하는 것이 중시됨에 따라, 주요 선진 국가들은 잠재적인 기술기회 발굴을 통하여 기술개발의 방향을 설정하거나 부가가치가 높은 융합 기술영역을 발굴하는 등의 활동을 통하여, 연구개발의 포트폴리오를 구축하기 위해 많은 노력을 하고 있다(이세준 외, 2008).

기술기회 발굴에 대한 방법론은 1990년 초반 Alan Porter를 중심으로 시작되었으며(김창석 외, 2013), 방법론적인 측면에서 볼 때 기술기회 발굴은 크게 전문가의 지식을 바탕으로 한 정성적인 평가 방법과 서지정보를 활용한 정량적 평가 방법으로 구분되어진다. 기술기회를 발굴하기 위한 과거의

연구는 주로 전문가의 지식을 바탕으로 한 전문가 인터뷰, 설문조사 등의 정성적 평가 방법이 주를 이루고 있었으며, 2000년대 이후부터는 정보기술의 발달에 의해 기술정보의 검색과 활용이 용이해짐에 따라 특허데이터 분석을 기반으로 한 서지학적 분석 방법론들이 많은 연구자들에 의해 다양하게 제시되었으며, 특히 특허 데이터베이스를 활용한 서지학적 분석 방법론들이 제시됨에 따라 기술기회 발굴에 대한 연구가 촉진 되었다(Choi et al, 2011). 이러한 특허 데이터베이스를 기반으로 한 서지학적 기술기회 발굴 방법론에 의해 전문가의 주관적 의견에 대한 의존성을 줄일 수 있게 되었고, 뿐만 아니라 기술기회 발굴을 위한 시간과 비용이 절약 되었으며, 예상치 못한 새로운 기술기회를 발굴 할 수 있는 가능성 또한 증가하였다(Yoon et al, 2011).

1.2 연구목표

정보기술의 발달에 힘입어 특허데이터를 기반으로 한 서지학적 분석 방법론들이 다양하게 제시됨으로써 기존의 전문가의 의견을 기반으로 한 기술기회 발굴연구의 단점을 보완 하였으며, 시간과 비용의 절약이나 예상치 못한 기술기회를 발굴 할 수 있었다. 하지만 이러한 장점들이 존재 함에도 불구하고 기존의 연구들은 다음과 같은 한계점을 지니고 있었다.

먼저 첫 번째로, 대부분의 기술기회 발굴에 관한 연구들은 기술기회를 발굴하고자 하는 대상의 내부 역량을 고려하지 못하고 있다(박현석 외, 2014). 국가적 차원에서 국가의 미래기술 예측을 위한 연구의 경우 유망기술을 선정하는 수준에서 의의를 찾을 수 있지만, 한 개인이나 기업 단위에

서 기술기회 발굴연구를 시행할 경우 연구 대상 개인이나 기업의 보유역량을 고려하지 못하고 기술기회를 예측한다면, 연구의 활용도가 떨어지고 실제 기술을 사업화 할 수 있는 가능성 또한 줄어들게 되어 연구의 가치가 떨어지게 된다.

둘째로, 특허분석을 기반으로 하여 기술기회를 발굴하는 연구들은 주로 특허의 서지정보인 IPC 특허분류코드, 인용정보 등을 활용하기 때문에(서원철 외, 2014) 특허문서가 가지고 있는 기술의 상세 내용을 알기 어렵다는 한계점을 가지고 있었다.

이에 본 연구에서는 한계점들을 극복하기 위해 다음과 같은 연구목표를 기반으로 한 연구방법론을 제시하고자 한다. 첫째, 기존의 연구들이 기업의 내부역량을 고려하지 못한 채 기술기회를 발굴했다는 한계점을 극복하기 위해 기술기회를 발굴하고자 하는 대상의 기술력을 가장 잘 설명하고 있는 특허데이터를 수집하여 내부역량을 파악한다. 둘째, 특허가 포함하고 있는 기술의 상세 내용을 파악 하기위해 텍스트 마이닝을 활용하여 접근 가능한 기술기회로써 기술테마를 발굴하고 추천한다.

1.3 연구프로세스

본 연구는 기술기회를 발굴하고자 하는 대상의 특허를 바탕으로 내부역량을 고려하여 실제 기업의 M&A, 사업영역확장 이나 공공연구기관과 정부의 R&D, 정책의 의사결정에 실제로 활용되어질 수 있는 기술테마를 추천하는 방법론을 제시하는 연구로써 연구의 프로세스는 다음과 같다.

먼저, 2장에서는 본 연구의 바탕이 되는 기술기회 탐색에 관한 기존연구의 방법론과 한계점등에 대해 설명하며, 본 연구에서 제시하는 방법론의 이론적 배경이 되는 Latent Dirichlet Allocation과 Association Rule Mining에 대해 설명 한다.

3장에서는 본 연구에서 제시하고 있는 연구 방법론에 대해 설명하기 위하여 기업의 내부역량을 파악하는 것을 시작으로 기술기회를 발굴하기까지의 전체적인 연구 흐름과 단계별 사용되는 기법 및 데이터, 툴 등에 대한 설명을 한다.

그리고 4장에서는 본 연구에서 제시하고 있는 방법론의 이해를 돕기 위해 3장에서 제시된 연구방법론의 과정에 대한 설명과 더불어, 방법론의 타당성을 검증하기 위해 대상 업체를 임의로 선정하여 사례연구를 진행하였다.

마지막 5장은 결론으로써, 본 연구에서 제시하고 있는 방법론의 의의 및 기여점과 한계점에 대한 설명과 더불어 한계점을 해결하기 위한 추후연구 방향을 제시하고 있다.

2.배경연구

2.1 기술기회 발굴

기술기회 발굴이란 앞서 서론에서 정의한 바와 같이 ‘기술 및 제품의 개발과 활용을 통해 수익 창출이 가능한 기회를 발굴하는 과정’으로써, 민간기업, 공공연구기관 및 국가적 차원에서 유망한 기술기회를 발굴하는 것은 글로벌 기술경쟁 속에서 지속적인 성장을 위한 필수적인 조건으로 인식되어지고 있으며, 기술기회를 발굴함으로써 민간차원에서 기업은 사업영역을 확장할 수 있고, 국가적 차원에서는 정책수립 및 기술관리, R&D등의 의사결정에 도움을 줄 수 있다(박현석 외, 2014).

기술기회 발굴에 대한 연구는 과거에는 전문 의견 기반의 정성적 평가 방법이 주를 이루었다. 이와 관련된 연구로는 델파이 분석 기법과 AHP분석 기법을 활용하여 유망기술을 도출한 후 우선순위 평가모형에 의해 기술을 예측하는 연구(김정석 외, 2009), 기술 인텔리전스를 활용하여 기술기회를 탐색하는 연구(윤병운 외, 2011) 등이 있었다. 하지만 이러한 연구들은 주로 전문가의 지식과 경험에 의한 주관적 의견이 포함되기 때문에 실제 공공연구기관이나 민간기업이 활용하기는 어려웠다.

정보기술의 발달과 함께 이러한 한계점들을 극복하기 위해 특허 데이터를 활용한 서지학적 분석 방법을 통한 정량적 분석 방법들이 많은 연구자들에 의해 제안되었다. 이러한 연구로는 특허 데이터에서 키워드를 추출하여

평균 빈도수와 증가율을 고려하여 기술기회를 발굴하는 연구(Yoon et al, 2012), 특허문서의 키워드를 기반으로 특허맵을 작성하여 기술기회를 발견하는 연구(Lee et al, 2009) 등이 있었다. 또한 특허를 활용하여 기술기회를 분석하는 다양한 연구 중 특허들 간의 인용관계를 정리하여 기술의 흐름과 확산 등의 연관성을 알 수 있는 연구가 제안 되었으며, 이와 관련한 연구로는 특허의 인용정보를 협업 필터링에 적용하여 보유기술에 기반한 기술 기회 탐색 연구(박영진 외, 2015) 등이 있었다.

이와 같이 특허문서를 활용한 기술기회 발굴연구는 지속적으로 발전되어지며 제시 되어왔으며, 산업의 기술 트렌드를 파악하고 기술기회가 발생하는 산업 분야를 파악하여 국가적 정책수립 및 기업의 기술 트렌드 변화에 대한 대응책 수립에 있어 많은 도움이 될 수 있었다. 하지만, 기술기회 발굴의 결과가 IPC특허분류코드와 같은 대분류의 범주이기 때문에 기술의 상세내용을 알기 어려워 실제 기업이나 공공 연구기관에서 활용하기에는 어려웠다. IPC란 스트라스부르 협정(Strasbourg Agreement Concerning the International Patent Classification)에 의하여 마련된 특허나 실용신안에 대한 통일된 분류 체계로써, 표 1.1과 같이 분 IPC 섹션을 나타내는 문자(예: G = 물리학)와 IPC 클래스를 나타내는 숫자(예: G02 = 광학), IPC 서브 클래스를 나타내는 문자(예: G02B = 광학요소), IPC 메인 그룹을 나타내는 숫자(예: G02B 006 = 라이트가이드), IPC 서브 그룹을 나타내고 있는 빗금 "/" 뒤의 숫자(G02B 006/44 = 광 전송 케이블)로 구성되어 있다. 이처럼 IPC특허 분류 코드는 섹션에서 서브그룹까지 약 70,000여개의 코드로 구분되어 지지만, 특허분석 등의 정량적 분석 등에서 활용될 때는 분류의 단위가 너무 세분화 되어 있어, 서브클래스 단위의 분류 650여개 코드로 모든 특허 데이터를 구분하고 있다.

표2.1 IPC특허분류코드

구분	섹션	클래스	서브클래스	메인그룹	서브그룹
표기	A~H	숫자 (2자리)	알파벳대문자 (1개)	숫자 (1~3개)	숫자 (2자리이상)
예시	G (물리학)	02 (광학)	B (광학요소)	006 (라이트가이드)	44 (광 전송 케이블)

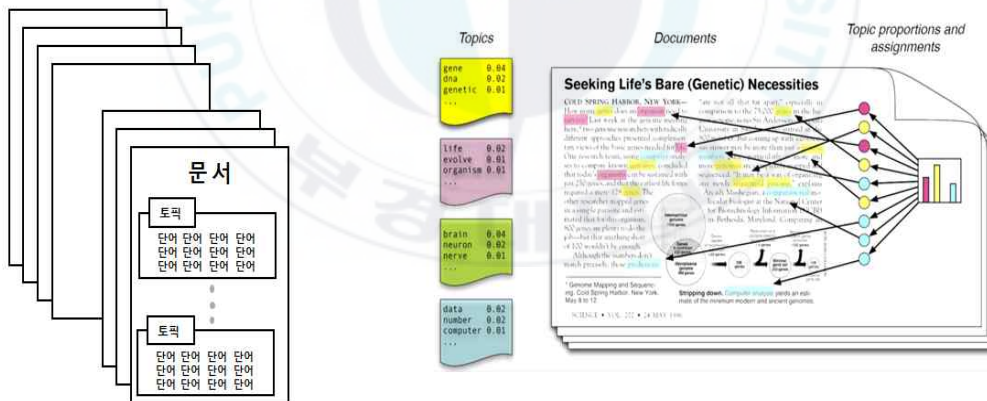
이에 많은 연구자들이 특허 내부에 존재하는 기술적 의미를 활용하여 기술기회를 발굴하기 위하여 텍스트 마이닝을 활용한 기술기회 발굴의 방법론을 다양하게 제시하였으며, 이와 관련된 연구로는 텍스트 마이닝 기법을 활용하여 서지학적 정보와 기술적 정보를 추출하여 활용하는 연구(Yoon et al, 2011), 토픽모델링을 활용하여 발명의 컨셉을 추출하고 그 변화를 모니터링하여 기술기회를 분석하는 연구(최하영 외, 2016) 등이 있었다.

따라서 본 연구에서는 기업의 내부역량을 고려한 기술기회를 발굴하기 위해서 텍스트 마이닝의 기법중 하나인 토픽모델링의 LDA기법을 활용하여 기술기회를 발굴하고자 하는 대상의 특허데이터가 가지고 있는 과학기술 정보를 기술테마의 단위로써 세분화 시켜 기술수준에 대한 정보를 얻고, ARM기법을 활용하여 내부역량과 강한 연관관계를 가지고 있는 잠재적인 기술테마를 추천함으로써, 추천받은 기술을 보다 상세히 알 수 있고, 공공기관이나 민간 기업의 연구개발 및 사업영역확장, M&A의 의사결정에 있어서 실제 활용이 가능한 방법론을 제시하고자 한다.

2.2 LDA(Latent Dirichlet Allocation)

토픽 모델링은 주어진 문서가 잠재적으로 가지고 있는 토픽(Topic)들을 추론하기 위한 확률적 모델로써, 문서내의 단어의 출현 빈도와 연관성을 고려하여 유사한 의미를 가진 단어들을 클러스터링 하는 방식으로 토픽을 추론한다(Steyvers et al, 2007). 토픽 모델링의 기법 중 가장 널리 쓰이는 모델은 Blei등에 의해 제시된 LDA로써, Dirichlet 분포를 이용하여 문서 내에서 출현하는 단어들이 특정한 토픽에 포함될 확률을 계산하고 도출된 결과를 통해 토픽을 구성하는 단어들의 집합을 추출하는 알고리즘이다(Blei et al, 2003).

그림2.1 Latent Dirichlet Allocation

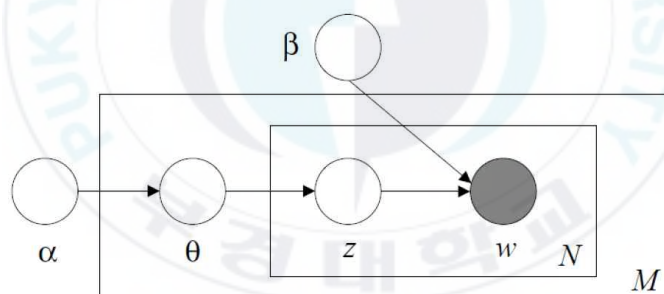


LDA모델의 핵심은 각각의 문서는 여러 토픽의 집합으로 구성되어있다고 가정하고, 그 문서를 이루는 단어들이 어떤 토픽과 연관되어있는지를 결정하는 것이다(Blei et al, 2003). LDA는 다수의 문서들로 이루어진 코퍼스(corpus)에서 문서가 어떤 내용으로 구성되어져 있는지 알 수 없을 때 잠

재되어 있는 토픽과 그 토픽을 구성하고 있는 단어들을 발췌하기 위하여 사용되어진다(조은경 외, 2014). LDA는 우선 임의의 토픽 개수를 정하고, 각 토픽을 구성하는 단어의 개수를 정한 후, 단어들을 임의 추출하여 어떤 토픽에 할당하면서 그 토픽에 적합하게 할당된 것인지를 추정하는 과정을 반복적으로 수행한다.

LDA는 그림1과 같은 도식으로 표현되어지며, 여기서 w 는 단어가 토픽에 포함될 확률, z 는 문서를 구성하고 있는 단어가 토픽에 포함될 확률, θ 는 문서가 토픽에 포함될 확률을 의미하며, α 는 θ 의 파라미터이고, β 는 w 의 파라미터, N 은 단어의 수, M 은 문서의 수를 의미한다.

그림 2.2 LDA 도식화



LDA는 주로 대량의 텍스트 데이터를 기반으로 정보를 검색해야 하는 문서의 주제 분류, 문서 간 유사도 계산 등을 위해 주로 사용되어져 왔으며(Krestel et al, 1995), 최근에는 LDA를 활용하여 특허데이터, 논문 등을 분석하여 과학, 기술과 관련된 트렌드와 이슈를 파악하는 연구가 활발히 진행 되었다(Tang et al, 2012). 이와 관련된 연구로는 토픽모델링과 감성 분석을 활용하여 제품 개발 기회를 도출하는 연구(정병기 외, 2016), 토픽

모델링의 결과로 나온 토픽이 각 특허문서에 포함될 확률을 활용하여 기술 기여도를 도출한 후 특허경쟁정보를 분석한 연구 등이 있었다(정병기 외, 2015).

본 연구에서는 기존 연구들이 내부역량을 고려하지 못하였고, 기술기회 발굴의 결과가 IPC 특허분류코드와 같이 세부적이지 못한 범주의 단위였다는 한계점을 해결하기 위해 기술기회를 발굴하고자 하는 대상의 특허데이터를 기반으로 LDA를 수행하여 내부역량을 고려한 토픽을 추출하고 있다. 이때, 특허는 과학기술에 대한 정보가 포함된 문서이므로, 특허 데이터를 기반으로 LDA를 수행한 결과 토픽들은 과학기술에 대한 정보를 포함하고 있을 것이다. 따라서 본 연구에서는 특허 데이터를 기반으로 추출된 토픽을 ‘기술테마’라 정의하고 있다.

본 연구에서는 LDA를 수행하기 위해, R 프로그래밍 패키지인 topicmodels를 활용하였으며, 그 외에도 LDA를 수행하기 위해 공개된 오픈소스 툴이 여러 가지가 있다. 대표적으로는 LDA를 제안한 David M. Blei가 만들어서 배포한 lda-c가 있고, Xuan-Hieu Phan and Cam-Tu Nguyen이 만들어서 배포한 GibbsLDA++있으며, Radim Řehůřek가 만들어서 배포한 Gensim이 있다.

2.3 ARM (Association Rule Mining)

ARM은 데이터 마이닝 기법중 하나로써, 대규모 데이터베이스에 포함된 항목들의 연관 관계를 찾기 위한 방법론이다(Kim et al, 2011). 연관규칙을 찾아내는 것은 마케팅에서 널리 사용되며 장바구니 분석 이라고도 불리는데, 어떤 제품들이 함께 구매되는가를 알아내는 것을 목적으로 한다(shmueli et al, 2007). ARM은 Agrawal의 연구 ‘Mining association rules between sets of items in large databases’ 에서 신뢰도 및 지지도에 대한 개념과 함께 소개되었고(Agrawal et al, 1993), 이 개념의 실제 구현을 위한 Apriori 알고리즘은 (Agrawal et al, 1994)에서 제시 되었다.

연관규칙의 아이디어는 항목들 사이의 모든 가능한 규칙들을 ‘if-then’ 형식으로 열거한 후 가장 높은 연관 관계를 나타낸 항목들을 선택하는 것이다. ‘if’ 부분을 조건부라 하고, ‘then’ 부분을 결론부라 하며, 의미 있는 연관 규칙을 도출하기 위해서는 지지도(support), 신뢰도(confidence), 향상도(Lift)로 구성된 3가지 평가 항목에 의한 판단이 필요하다(서원철 외, 2014).

지지도는 단순히 조건부와 결론부 모두에 포함된 항목집합의 거래수를 나타내며, 도출된 결론의 유용성(utility)을 의미한다. 신뢰도는 조건부가 이미 출현 했을 때 결론부가 출현할 조건부 확률을 나타내며, 도출된 결론의 확실성(certainty)을 의미한다. 향상도는 해당 규칙의 신뢰도를 조건부와 결론부의 독립을 가정한 신뢰도로 나눈 값이며, 조건부와 결론부의 출현 간 연관관계를 나타내는 것으로써 도출된 규칙의 연관성(correlation)을 의

미한다. 향상도가 1보다 크면 양의 연관성을 가지고 있다고 판단 할 수 있으며, 의미 있는 연관규칙만을 추출하기 위해서는 지지도 및 신뢰도 값이 충분하게 존재해야 하고, 향상도는 1보다 큰 값을 나타내야만 한다. 표1은 3가지 지표에 대한 정의와 표현 식을 나타낸다(shmueli et al, 2007), (서원철 외, 2014).

표 2.2 ARM 지표

지표	설 명	표현 식
지지도	$A \rightarrow B$ 로 발견된 규칙의 유용성	$P(A \cap B)$
신뢰도	$A \rightarrow B$ 로 발견된 규칙의 확실성	$P(B A)$
향상도	$A \rightarrow B$ 로 발견된 규칙의 연관성	$\frac{P(B A)}{P(B)}$

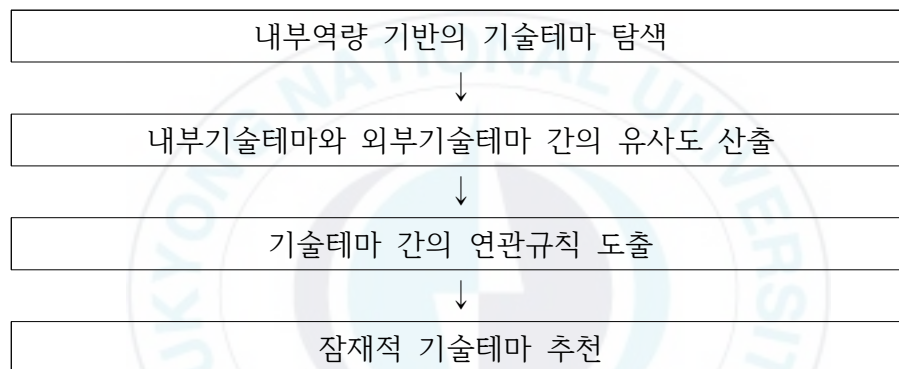
연관 규칙을 찾기 위해 많은 알고리즘 들이 제안되었지만, 그 중 가장 많이 쓰이는 알고리즘은 Apriori 알고리즘이다. Apriori 알고리즘은 단일 항목으로만 이루어진 규칙들을 생성하고, 빈도수 높은 항목들 중에서 지지도 값이 임계치 이상인 항목들을 찾아낸 후, 그 집합으로부터 임계치 이상의 신뢰도를 가지는 규칙들을 모두 찾아낸다. 이 후 단일 항목들 간에 생성된 규칙의 빈도가 높으면, 단일 항목을 부분 집합으로 포함하는 여러 개의 항목 간의 규칙 빈도수가 단일 항목의 규칙 빈도수 보다 높을 수 없다는 아이디어에 의해 모든 크기의 규칙을 생성한다(shmueli et al, 2007). 예를 들어 $\{A\} \rightarrow \{B, C\}$ 의 규칙이 존재 할 때 이 규칙의 빈도수는 $\{A\} \rightarrow \{B\}$ 규칙의 빈도수 보다 높을 수 없다. 이는 하나의 규칙이 주어졌을 때 새로운 항목을 추가하게 되면 지지도는 전보다 좋아질 수 없다는 것을 의미한다.

이와 같이 항목들 간의 강한 연관관계를 가지는 규칙을 찾기 위한 ARM 기법은 기술기회를 발굴하기 위한 연구에서도 다양하게 활용되어져 왔으며, 이와 관련된 연구로는 등록된 특허가 가지는 IPC특허분류 코드들 간의 연관규칙을 통해 융합패턴을 분석하여 융합 기술기회를 발굴하기 위한 연구(최하영 외, 2015), 기업이 보유하고 있는 제품 포트폴리오를 기반으로 연관규칙을 발견하여 제품 기술기회를 발굴하는 연구(서원철 외, 2014) 등이 있었다. 하지만 이러한 연구들은 앞서 명시한 기존 연구의 한계점과 마찬가지로 연구의 결과로써 발굴된 기술기회의 상세내용을 알기 힘들어 활용하기 어렵다는 한계점을 가지고 있었으며, 본 연구에서는 이를 해결하기 위해 LDA의해 추출된 기술테마들 간의 의미 있는 연관규칙을 통해 새로운 기술기회를 발굴하고자 한다.

3.연구 방법론

본 연구는 기업의 내부역량을 고려한 후, 새로운 기술기회를 탐색하기 위한 방법론으로써, 그림3.1과 같은 연구 방향을 따른다.

그림 3.1 연구프로세스



3.1 내부역량 기반의 기술테마 탐색

3.1.1 데이터 수집 및 내부기술테마 도출

본 연구는 기존의 기술기회 발굴 연구가 내부 역량을 고려하지 못했다는 한계점을 극복하기 위하여 내부 역량을 고려하는 과정을 가장 먼저 실시한다. 내부역량을 고려하기 위해 우선 대상기업을 선정하고, 대상 기업의 기술력을 가장 쉽고 정확하게 파악할 수 있는 특허데이터를 수집하는 과정이 필요하다. 본 연구에서는 선정된 대상기업의 특허데이터를 수집하기 위해

Wips-on(<http://www.wipson.com>) 특허 검색 사이트를 활용하였다.

이후 내부역량을 기술테마 단위로 분석하기 위해 LDA를 실시하여 특허 데이터의 토픽을 발췌 한다. 이때 수집된 특허 문서에는 대상기업이 출원하고자 하는 기술에 대한 설명과 정보가 명시 되어져 있기 때문에, LDA결과 토픽들 또한 해당 과학 기술과 연관성이 큰 단어들로 구성되어져 있을 것이다. 따라서 본 연구에서는 LDA에 의해 추출된 토픽을 보유기술에 대한 정보를 가장 잘 보여주고 있는 ‘내부 기술테마’라 정의 하고 있다. LDA를 수행하기 위해 본 연구에서는 R programing의 topicmodels패키지를 활용하였으며, 특허문서의 요약문만을 발췌하여 코퍼스를 구성하였고, 토픽의 수는 80개 각각의 토픽을 구성하는 단어의 수는 10개로 설정하였다.

3.1.2 외부기술테마 도출

본 과정은 내부기술테마를 추출한 이후, 이를 기반으로 추천받을 외부기술테마를 도출하는 과정으로, 외부기술테마 역시 특허데이터를 기반으로 LDA과정을 수행하여 얻어질 결과이기 때문에 추천받을 기술테마에 대한 특허데이터를 수집하는 과정이 우선적으로 필요하며, 이는 앞선 내부기술테마 도출 결과를 활용한다. 내부기술테마는 LDA의 결과로 80개의 토픽이 각각 10개의 단어로 구성되어져 있는 형태이다. 본 연구에서는 표3.1과 같이 10개의 단어중 가중치가 높은 상위 5개의 단어를 기반으로 특허검색을 위한 쿼리문을 만들고 이를 바탕으로 Wips-on에서 외부 특허데이터를 수집하였다. 이때, 상위 5개의 가중치는 단순히 절반의 단어로 쿼리를 구성한 것이며, 쿼리를 구성하는 단어 하나의 증감에 따라 수집되는 특허데이터 개수의 증감이 크기 때문에 추후 연구에서는 쿼리문을 구성하는 단어의 개

수를 보다 객관적으로 정하는 과정이 필요하다. 본 연구의 LDA과정에 있어 불용어가 완벽하게 제거되어지지 못해 의미 없는 단어들이 포함되어 가중치를 기반으로 임계치를 설정하지 못하였다.

표 3.1 외부기술 탐색을 위한 쿼리문 작성 예시

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
materials	battery	high	monomer	electrolyte
application	cells	due	resin	secondary
thermoelectric	module	conductivity	copolymer	battery
conversion	modules	mechanical	vinyl	lithium
plastic	mounted	ionic	latex	nonaqueous
devices	pack	simple	impact	solvent
films	plurality	strength	rubber	salt
substrates	side	easy	aromatic	solution
optoelectronic	case	low	thermoplastic	characteristics
substrate	members	solid	graft	additive
쿼리				
(materials and application and thermoelectric and conversion and plastic) or (battery and cells and module and modules and mounted) or (high and due and conductivity and mechanical and ionic) or (monomer and resin and copolymer and vinyl and latex) or (electrolyte and secondary and battery and lithium and nonaqueous)				

이후 외부기술테마를 도출하기 위해 표3.1과 같이 작성된 쿼리문을 토대로 특허데이터를 수집하고, 특허의 요약문만을 발췌하여 다시 한 번 코퍼스를 만들어 이를 바탕으로 LDA를 수행하여 토픽을 추출하였다. 이 과정도 앞선 과정과 마찬가지로 토픽의 수는 80개, 이를 구성하는 단어의 수는 10개로 지정하였으며, 이결과로 추출된 토픽은 수집된 외부특허데이터를 기반으로 발췌된 토픽이므로, 본 연구에서는 ‘외부 기술테마’라 정의하고 있다.

3.2 내부기술테마와 외부기술테마 간의 유사도 산출

본 연구에서는 외부기술테마들 간의 연관성분석을 통하여 잠재적인 기술테마를 추천하기 위한 네트워크를 구성한다. 하지만 단순히 외부기술테마들 간의 연관성분석을 통하여 추천 네트워크를 구성하는 것은 기존의 내부역량을 고려하지 않은 기술기회 발굴연구와 차이가 없기 때문에 본 연구는 외부기술테마와 내부기술테마 사이의 유사도를 산출하여 연관성 분석의 조건부 부분에서 내부기술테마와 외부기술테마가 유사하다면, 이를 내부기술테마로 간주하고, 결론부 부분에서 내부기술테마와 외부기술테마가 유사하다면 이미 보유한 기술테마이기 때문에 추천받을 필요가 없어, 생략하는 과정을 수행한다.

본 연구에서 유사도 산출은 python library Genism을 활용하여 산출했으며, Gensim은 cosine 유사도를 기반으로 유사도를 산출하지만, 단순히 단어와 단어가 같은지 만을 고려하는 것이 아니라 LDA결과 토픽을 구성하고 있는 단어들 간의 연관성 까지 고려하여 유사도를 산출해 준다. 그 결과 외부기술테마 하나에 대한 모든 내부기술테마의 유사도 값이 산출되고, 본 연구에서는 그중 임계치 이상의 유사도 값을 가져야 유사하다고 정의하고 있다.

3.3 기술테마 간의 연관규칙 도출

3.3.1 문서-기술테마 매트릭스

ARM은 항목들 간의 동시출현 확률에 기반 하여 의미 있는 규칙을 탐색하는 것이므로, 연관규칙을 도출하기 위해서는 우선적으로 하나의 기술테마가 하나의 특허문서 내에서 어떠한 다른 기술테마와 동시에 출현하였는지를 알아야한다. 본 연구에서 정의하고 있는 기술테마는 LDA과정에서 추출된 토픽이기 때문에, 기술테마들 간의 동시출현 빈도를 알기위하여 LDA결과로 알 수 있는 문서별 각 토픽의 가중치 값을 활용한다.

하지만 토픽의 가중치 값은 단순히 토픽을 구성하는 단어가 특허문서에 포함되었는지 만을 고려한 것이 아니라 토픽을 구성하는 단어들 간의 연관성에도 영향을 받기 때문에 하나의 특허문서에 대해 모든 토픽들이 아주 미미한 값이라도 가중치를 가지고 있다. 그렇기 때문에 이 값을 토대로 ARM을 수행하게 되면, 모든 기술테마들은 하나의 특허문서 내에서 모두 동시에 출현하여 모든 기술테마들 간에 유의미한 규칙이 존재한다는 결론이 도출되어지게 된다. 이에 본 연구에서는 하나의 특허문서에서 상위X개의 가중치 값을 가지는 기술테마들 만을 유의미한 값으로 정의하는 작업을 통해 하나의 특허 문서에는 X개의 기술테마들만이 동시에 출현하게 하여 문서-기술테마 매트릭스를 재구성하였다.

3.3.2 기술테마 간의 연관규칙 도출

본 과정은 추천 가능한 잠재적인 외부기술테마를 찾기 위한 과정으로써, 이전 과정에서 도출한 문서-외부기술테마 매트릭스를 활용하여 ARM 기법을 적용하는 단계이다. 본 연구에서는 ARM을 수행하기 위하여 R프로그래밍 arules 패키지를 활용하였다.

ARM은 조건부와 결론부 모두에 포함된 항목집합의 거래수를 나타내는 지지도와, 조건부가 이미 출현했을 때 결론부가 출현할 조건부 확률을 나타내는 신뢰도, 해당 규칙의 신뢰도를 조건부와 결론부의 독립을 가정한 신뢰도로 나눈 값인 향상도 총 3가지 평가 항목에 의해 평가되며, 유의미한 규칙만을 찾기 위해 지지도와 신뢰도에 각각의 임계치를 설정하여 임계치 이상의 값을 가져야하고, 향상도의 경우 양의 상관관계만을 고려하기 위해 향상도가 1이상의 값을 가져야 한다고 정의한다.

3.4 잠재적 기술테마 추천

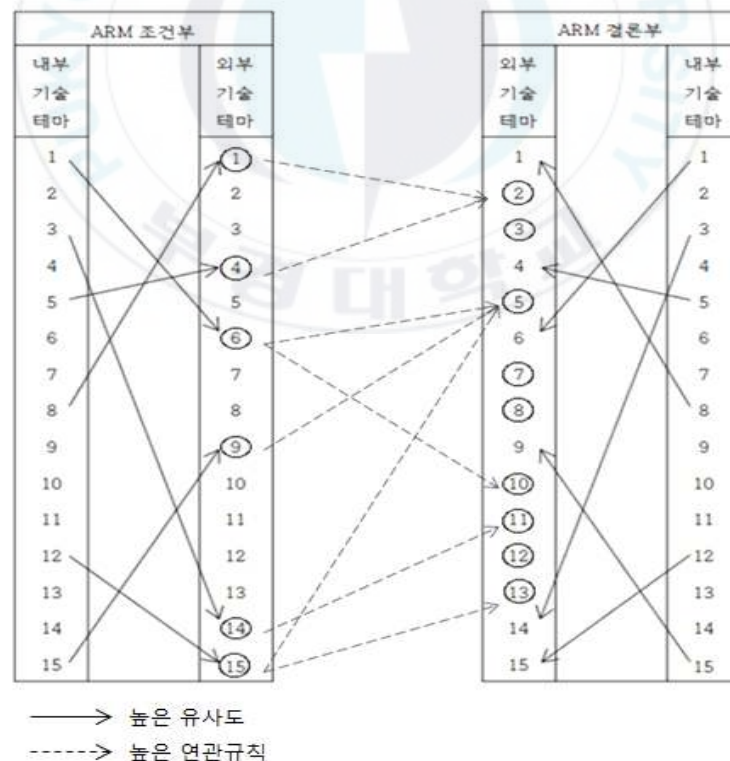
3.4.1 최종추천네트워크

본 과정은 앞선 과정의 결과 데이터를 종합하여 최종적인 잠재 추천 기술테마를 도출하는 과정이다. 본 연구는 보유역량을 바탕으로 기술기회를 발굴하기 위한 연구로써, 최종 추천 네트워크는 보유기술에서 보유하지 않은 기술테마를 추천 받는 형태로 구성되어 져야 한다. 이를 위해 앞선 3.3.2의 ARM 결과의 조건부 항목에서 3.2의 유사도가 높은 항목만을 필터링 한

후, ARM 결과의 결론부 항목에서 유사도가 낮은 항목만을 필터링 하는 작업을 수행한다.

이는 외부기술테마 간의 연관규칙을 도출하여 의미 있는 규칙을 찾은 후, 조건부의 외부기술테마와 내부기술테마 사이의 유사도가 높을 경우 조건부의 외부기술테마는 내부기술로 간주하고, 결론부의 외부기술테마와 내부기술테마 사이의 유사도가 높을 경우 최종 결과에서 제외함으로써, 내부기술테마와 외부기술테마간의 의미 있는 규칙을 찾기 위한 것이다. 그림3.2는 추천 네트워크 구성을 도식화 한 것이다.

그림 3.2 추천 네트워크 구성



3.4.2 결과해석

앞 선 3.3.1 ~ 3.4.1의 과정을 통해 도출된 추천 네트워크를 통해 잠재적으로 추천 받을 수 있는 외부기술테마 군을 정의 하였다. 본 과정에서는 최종 추천 네트워크의 조건부와 결론부의 테마들이 LDA 결과로 도출된 10개의 단어로 구성되어 있기 때문에, 이 단어들의 의미와 연관관계를 토대로 테마의 명칭을 정하는 과정이 필요하며, 정해진 테마의 이름을 토대로 조건부와 결론부의 테마들이 타당한 근거에 의해 추천 되었는지 탐구해보아야 한다.

3.4.3 상세특허추천

지금까지 본 연구를 통해서 내부 역량을 고려한 기술기회 발굴의 결과로써 기술테마를 제시하였다. 하지만 본 연구의 결과가 기업의 M&A나 사업영역확장 등에서 실용적으로 사용되기 위해서는 보다 상세한 기술 정보를 필요로 할 수 있다. 특허는 발명과 관련된 구체적인 기술 및 과학 정보를 포함하고 있는 문서로써(노현정 외, 2009), 특허 데이터베이스에는 세상에 존재하는 기술정보의 90% 이상이 담겨져 있다고 알려져 있다(Zha et al, 2010). 그렇기 때문에 추가적인 연구로써, 앞선 3.3.1의 문서-기술테마 매트릭스의 가중치를 활용하여 추천된 기술테마의 가중치가 높은 특허문서를 역추적 한 후, 특허 리스트를 도출하여 제시 한다.

4. 사례연구

본 연구에서는 연구의 타당성을 검증하기 위해 사례연구를 실시한다.

4.1 내부역량 기반의 기술테마 탐색

4.1.1 데이터수집 및 내부기술테마 도출

본 과정은 기업의 기술력을 파악한 후, 이를 바탕으로 내부기술테마를 도출하는 과정으로써, 이를 위해 우선 사례연구를 위한 대상 기업을 선정 하였다. 대상 기업으로는 최근 화두가 되고 있는 전기 자동차와 관련하여 2차 전지 사업에서 두각을 드러내고 있는 'LG화학'(출원인 코드:UN000021)을 선정하였으며, 특허 검색 사이트 Wips-on을 통하여 2010년~2015년 까지 미국에 등록되어진 1,726건의 LG화학 특허 데이터를 수집하였다.

이후 내부기술테마를 도출하기 위하여 1,726건의 특허 데이터 요약문을 토대로 코퍼스를 구성한 후 LDA를 실시하였으며, 본 연구에서 LDA를 수행하기 위해 R 프로그래밍의 topicmodels 패키지를 활용하였고 LDA의 전처리 과정인 불용어 처리와 topicmodels 패키지에서 LDA를 수행함에 있어 필요로 하는 Documnet-Term matrix 형태를 구성하기 위해 텍스트마이닝 패키지인 tm 패키지를 활용 하였다. 또한 특허 문서의 특성상 빈번하게 등장하는 단어들(Invention, product, provide, suggest, first, second 등)을 LDA 결과에서 제거하는 과정을 통해 특허에서 명시하고 있는 기술과 관

련된 단어로 구성된 토픽을 발췌 하였다.

LDA의 결과 부록1과 같이 토픽 80개가 도출 되었으며, 본 연구에서는 이와 같이 LDA결과로 도출된 토픽을 내부기술테마라 정의하였다. 부록1의 결과는 LG화학의 특허 요약문을 토대로 구성된 코퍼스를 기반으로 한 LDA 결과이기 때문에 주로 화학과 관련된 단어들이 토픽을 구성하고 있다는 것을 알 수 있으며, 2010~2015년 사이 2차 전지와 관련된 기술에 많은 중점을 두고 있었기 때문에 이와 관련된 단어들도 많이 발췌 된 것을 알 수 있다. 표4.1은 LDA결과로 도출된 내부기술테마의 일부이다.

표 4.1 내부기술테마(일부)

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
materials	battery	high	monomer	electrolyte
application	cells	due	resin	secondary
thermoelectric	module	conductivity	copolymer	battery
conversion	modules	mechanical	vinyl	lithium
plastic	mounted	ionic	latex	nonaqueous
devices	pack	simple	impact	solvent
films	plurality	strength	rubber	salt
substrates	side	easy	aromatic	solution
optoelectronic	case	low	thermoplastic	characteristics
substrate	members	solid	graft	additive

4.1.2 외부기술테마 도출

본 과정은 내부기술과 연관성이 높은 특허를 기반으로 다시 한 번 LDA를 수행하여 외부기술테마를 도출하는 과정으로, 먼저 내부기술과 연관성이 높은 특허를 찾기 위해 내부기술테마를 구성하는 10개의 단어 중 상위 5개 단어의 조합을 통해 특허 검색 쿼리문을 작성하였다. 그림 4.1은 최종적으로 구성된 쿼리문이다.

그림 4.1 외부기술 탐색을 위한 쿼리문

(materials and application and thermoelectric and conversion and plastic) or (battery and cells and module and modules and mounted) or (high and due and conductivity and mechanical and ionic) or (monomer and resin and copolymer and vinyl and latex) or (electrolyte and secondary and battery and lithium and nonaqueous) or (column and distillation and wall and conventional and dividing) or (acid and methacrylic and step and solution and gas) or (electrochemical and stability and thermal and mixture and separation) or (water and phosphate and iron and solution and controller) or (carbon and core and component and crystalline and coating) or (resin and composition and photosensitive and resistance and relates) or (optical and light and filter and element and image) or (properties and prepared and low and compounds and superior) or (agent and flame and retardant and styrene and groups) or (temperature and high and heating and discharging and charging) or (preparing and capable and lactate and gene and pla) or (surface and area and transparent and opening and ratio) or (manufacturing and reduced and manufactured and article and improve) or (type and time and driving and mandrel and controlled) or (electric and disclosed and insulation and hybrid and pouch) or (battery and case and part and secondary and contact) or (portion and sealing and portions and tape and attached) or (conductive and energy and level and ntype and ptype) or (negative and retardation and positive and index and relates) or (member and assembly and portion and disposed and coupled) or (compound and comprising and aliphatic and nitrile and residual) or (bus and bar and wire and bars and laser) or (lower and upper and frame and member and parts) or (crystal and liquid and display and alignment and film) or (battery and connection and member and terminal and cell) or (membrane and electrolyte and fuel and ion and reverse) or (safety and cap and assembly and top and gasket) or (disclosed and resistance and isolation and fault and function) or (reaction and reactor and unsaturated and producing and mixture) or (group and formula and compound and represented and consisting) or (adhesive and excellent and acrylic and property and relates) or (coating and porous and particles and binder and inorganic) or (unit and units and plurality and number and backlight) or (disclosed and molding and capable and mold and effectively) or (battery and pack and total and slave and management) or (specific and panel and artificial and surface and composite) or (catalyst and olefin and preparing and relates and high) or (gas and potential and comprises and generation and disclosed) or (weight and parts and molecular and based and comprising) or (battery and cooling and module and heat and channel) or (electrode and active and secondary and layer and outer) or (amount and groups and disclosed and viscosity and trihalosilane) or (plate and base and polarizing and plates and side) or (film and polarizer and optical

and thin and angle) or (pattern and substrate and ink and line and forming) or (preparing and toner and paper and polymerized and improved) or (electrode and assembly and leads and lead and electrodes) or (metal and transition and complex and group and nanoparticles) or (organic and light and emitting and electronic and hole) or (body and main and housing and cover and edge) or (power and charge and control and capacity and signal) or (glass and molten and bath and float and capable) or (part and fixed and housing and cone and spinning) or (battery and voltage and estimating and resistance and soc) or (composition and addition and monomer and norbornene and solvent) or (battery and pack and air and vehicle and external) or (composition and excellent and resistance and curable and surface) or (voltage and battery and discharge and sensing and unit) or (cathode and anode and active and secondary and separator) or (sheet and disclosed and sheets and laminate and size) or (slurry and polishing and cmp and disclosed and silicon) or (copolymer and block and relates and olefin and repeating) or (coolant and port and inlet and space and flow) or (polymer and polymerization and monomer and photoreactive and preparing) or (heat and tube and exchanger and exchange and dissipation) or (comprising and disclosed and andor and reduction and shows) or (roll and board and excellent and metallic and paste) or (layer and layers and forming and lay and sequentially) or (comprises and comprising and relates and alkali and terephthalaldehyde) or (lithium and oxide and active and metal and transition) or (battery and system and time and vector and plurality) or (cell and battery and balancing and plurality and solar) or (voltage and circuit and microprocessor and level and signal) or (cerium and powder and oxide and preparing and precursor) or (particle and particles and diameter and range and size)

이후, Wips-on에서 그림4.1과 같이 작성된 쿼리문을 통해 2010년~2015년까지 미국에 등록된 특허 11,319건의 특허 데이터를 수집하였으며, 4.1.1의 과정과 마찬가지로 기술테마를 도출하기 위하여 특허 데이터의 요약문으로 구성된 코퍼스를 기반으로 LDA를 수행하였으며 그 결과는 부록2와 같다. 표4.2는 LDA결과로 도출된 외부기술테마의 일부이다.

표 4.2 외부기술테마(일부)

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
direction	electrode	block	portion	pressure
angle	active	copolymer	portions	flexible
axis	lithium	membrane	aperture	input
polarizing	battery	oxygen	disposed	techniques
polarizer	secondary	comprising	edge	computing
optical	positive	segment	receiving	higher
index	negative	improved	extending	hinge
parallel	electrolyte	hydrophobic	face	sensitive
polarization	collector	copolymers	holder	respective
retardation	nonaqueous	polyethylene	center	hydrocarbons

4.2 내부기술테마와 외부기술테마 간의 유사도 산출

본 연구의 최종적인 결론은 외부기술테마 간의 연관성 분석을 통해 도출되어진다. 하지만 앞서 설명한 바와 같이 단순히 외부기술테마 간의 연관성을 고려하여 결과를 도출하는 것은 기존의 연구와 마찬가지로 내부역량을 고려하지 못하기 때문에 내부기술테마와 외부기술테마 간의 유사도 산출과정을 통하여 조건부에서 외부기술테마 중 내부기술테마와 유사한 테마를 찾아 이를 내부기술테마로 정의하고, 결론부에서 외부기술테마 중 내부기술테마와 유사하지 않은 테마를 찾아 추천 가능한 후보군으로 정의하는 작업이 필요하다.

그림 4.2 유사도 산출 테이블

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1		lg1	lg2	lg3	lg4	lg5	lg6	lg7	lg8	lg9	lg10	lg11	lg12	lg13	lg14	lg15
2	us1	(0.00)	0.00	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)	(0.00)	0.00	0.00	0.19	0.00	(0.00)	0.00
3	us2		0.00	0.14	0.00	0.00	0.69	(0.00)	0.00	0.00	0.00	0.00	(0.00)	0.00	(0.00)	(0.00)
4	us3		(0.00)	0.00	(0.00)	0.30	(0.00)	0.00	(0.00)	(0.00)	0.00	(0.00)	(0.00)	0.00	0.00	(0.00)
5	us4		0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.00	0.00	0.00	0.00	0.00	(0.00)	0.00
6	us5		(0.00)	0.00	(0.00)	(0.00)	(0.00)	0.00	0.00	0.00	(0.00)	0.00	0.00	(0.00)	(0.00)	(0.00)
7	us6		0.00	(0.00)	(0.00)	(0.00)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.34
8	us7		0.00	0.00	0.41	(0.00)	0.20	(0.00)	(0.00)	0.20	(0.00)	0.20	0.00	(0.00)	0.20	0.00
9	us8		0.00	0.00	0.00	(0.00)	0.00	(0.00)	0.00	0.00	(0.00)	(0.00)	0.00	(0.00)	0.00	(0.00)
10	us9		0.33	0.00	(0.00)	0.00	0.00	(0.00)	(0.00)	0.00	0.00	(0.00)	0.00	0.16	(0.00)	0.00
11	us10		(0.00)	(0.00)	0.00	(0.00)	(0.00)	0.00	(0.00)	0.00	0.00	0.38	(0.00)	0.38	(0.00)	(0.00)
12	us11		0.37	(0.00)	0.00	0.00	(0.00)	0.73	0.00	(0.00)	(0.00)	0.00	0.00	(0.00)	0.00	0.00
13	us12		(0.00)	0.00	0.00	(0.00)	(0.00)	(0.00)	0.00	0.00	0.24	0.47	0.00	(0.00)	0.00	0.24
14	us13		(0.00)	0.00	0.00	(0.00)	(0.00)	0.00	(0.00)	0.00	0.00	0.44	0.00	(0.00)	0.00	0.00
15	us14		0.00	0.28	(0.00)	(0.00)	0.28	0.00	0.00	(0.00)	0.00	(0.00)	0.00	0.00	(0.00)	(0.00)
16	us15		0.00	0.00	0.00	0.00	0.00	0.00	(0.00)	0.00	(0.00)	0.00	0.00	(0.00)	(0.00)	0.00
17	us16		(0.00)	0.89	(0.00)	0.00	0.13	0.00	(0.00)	(0.00)	0.00	0.00	0.00	(0.00)	(0.00)	(0.00)
18	us17		0.00	0.00	(0.00)	0.00	(0.00)	0.00	0.00	0.00	(0.00)	(0.00)	(0.00)	0.97	(0.00)	0.00
19	us18		0.29	0.00	(0.00)	(0.00)	(0.00)	(0.00)	0.00	(0.00)	(0.00)	(0.00)	(0.00)	0.29	0.00	0.00
20	us19		(0.00)	(0.00)	(0.00)	(0.00)	0.00	0.00	0.00	0.00	0.00	(0.00)	(0.00)	(0.00)	0.00	(0.00)
21	us20		(0.00)	0.00	0.16	(0.00)	0.80	(0.00)	0.16	0.00	0.16	0.00	0.00	0.00	(0.00)	0.00
22	us21		(0.00)	0.00	(0.00)	(0.00)	(0.00)	(0.00)	0.00	0.00	0.55	0.28	(0.00)	(0.00)	(0.00)	0.00
23	us22		(0.00)	(0.00)	(0.00)	(0.00)	0.00	0.00	0.35	0.00	0.00	0.00	0.00	0.00	0.00	(0.00)
24	us23		0.61	(0.00)	0.00	0.00	0.00	0.00	0.00	0.00	(0.00)	(0.00)	0.00	(0.00)	0.00	(0.00)
25	us24		(0.00)	(0.00)	0.00	(0.00)	0.00	(0.00)	(0.00)	(0.00)	(0.00)	0.16	0.00	0.00	(0.00)	0.00
26	us25		0.00	0.00	0.00	(0.00)	0.00	(0.00)	(0.00)	0.00	(0.00)	(0.00)	0.00	0.00	0.00	(0.00)
27	us26		0.00	0.00	0.00	0.00	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)	0.00	0.00	0.00	(0.00)
28	us27		(0.00)	(0.00)	0.00	(0.00)	(0.00)	(0.00)	0.00	0.00	0.00	(0.00)	0.00	0.00	0.00	0.00
29	us28		0.38	0.00	0.00	0.00	0.00	(0.00)	0.00	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)	0.00
30	us29		0.00	0.00	(0.00)	(0.00)	0.00	0.00	0.00	(0.00)	(0.00)	0.00	0.00	(0.00)	(0.00)	(0.00)
31	us30		(0.00)	(0.00)	0.00	(0.00)	0.00	0.22	(0.00)	(0.00)	0.00	(0.00)	0.00	0.00	(0.00)	0.00

그림4.2는 Gensim의 유사도 산출기능을 통해 산출된 유사도산출결과이며

X축은 내부기술테마, Y축은 외부기술테마를 의미하고, 각 셀의 값들은 기술테마간의 유사도를 의미하고 있다. 이중 임계치 이상의 유사도 값을 가져야 유사하다고 정의하기 위해 임계치를 설정하였으며, 본 연구에서는 유사도의 평균값인 0.66을 소수 둘째 자리에서 올림 하여 0.7을 임계치로 설정하였다. 그 결과 외부기술테마 80개와 내부기술테마 80개의 모든경우의 수 6,400개중 33쌍의 기술테마들이 유사하다는 결과가 도출되었다. 표4.3은 유사도 산출결과이며, IN은 Internal technology themes를 의미 하고 EX는 External technology themes를 의미한다.

표 4.3 유사도 산출 결과

내부기술테마	외부기술테마	내부기술테마	외부기술테마
IN01	EX73	IN38	EX15
IN02	EX16	IN44	EX27
IN05	EX20	IN45	EX67
IN06	EX11	IN49	EX01
IN12	EX17	IN49	EX62
IN25	EX04	IN52	EX43
IN25	EX38	IN54	EX53
IN25	EX39	IN55	EX30
IN25	EX55	IN62	EX80
IN25	EX75	IN68	EX27
IN26	EX78	IN69	EX69
IN29	EX09	IN70	EX29
IN30	EX33	IN73	EX26
IN35	EX12	IN75	EX40
IN35	EX24	IN79	EX52
IN35	EX63	IN80	EX19
IN37	EX54		

4.3 기술테마 간의 연관규칙 도출

4.3.1 문서-기술테마 매트릭스

ARM과정을 진행하기 위해서는 하나의 기술테마가 하나의 특허 내에서 어떠한 다른 기술테마와 함께 출현하였는지를 알아야 한다. 본 연구에서 정의하고 있는 기술테마는 LDA의 결과인 토픽이므로, 한 문서에 대한 토픽의 가중치를 활용하여 기술테마 간의 동시 출현 항목을 알 수 있다. 그림 4.3은 문서-토픽의 가중치 매트릭스 일부이며, X축은 토픽의 번호, Y축은 문서의 번호를 의미한다.

그림 4.3 문서-토픽 가중치 매트릭스(일부)

#	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1																					
2		1	0.005435	0.005435	0.005435	0.222826	0.005435	0.005435	0.005435	0.01413	0.005435	0.005435	0.005435	0.005435	0.005435	0.005435	0.005435	0.005435	0.005435	0.005435	0.005435
3		2	0.007813	0.007813	0.007813	0.007813	0.020313	0.007813	0.007813	0.032813	0.007813	0.020313	0.007813	0.007813	0.007813	0.007813	0.007813	0.007813	0.007813	0.007813	0.007813
4		3	0.005531	0.005531	0.005531	0.005531	0.005531	0.005531	0.005531	0.005531	0.014381	0.005531	0.005531	0.005531	0.005531	0.005531	0.005531	0.005531	0.005531	0.005531	0.005531
5		4	0.006188	0.006188	0.006188	0.075495	0.006188	0.006188	0.006188	0.006188	0.006188	0.006188	0.006188	0.006188	0.016089	0.006188	0.006188	0.006188	0.045792	0.006188	0.006188
6		5	0.0125	0.004808	0.004808	0.0125	0.004808	0.004808	0.004808	0.004808	0.004808	0.004808	0.004808	0.004808	0.004808	0.004808	0.004808	0.004808	0.004808	0.004808	0.004808
7		6	0.004699	0.004699	0.004699	0.034774	0.004699	0.004699	0.012218	0.004699	0.004699	0.004699	0.004699	0.004699	0.012218	0.004699	0.004699	0.004699	0.004699	0.004699	0.004699
8		7	0.005208	0.005208	0.005208	0.030208	0.005208	0.005208	0.005208	0.013542	0.005208	0.005208	0.005208	0.005208	0.046875	0.013542	0.005208	0.013542	0.013542	0.005208	0.005208
9		8	0.010246	0.010246	0.010246	0.010246	0.010246	0.010246	0.010246	0.010246	0.010246	0.010246	0.010246	0.010246	0.010246	0.010246	0.010246	0.010246	0.010246	0.010246	0.010246
10		9	0.009921	0.009921	0.009921	0.009921	0.009921	0.009921	0.009921	0.009921	0.009921	0.009921	0.009921	0.009921	0.009921	0.009921	0.009921	0.009921	0.009921	0.009921	0.009921
11		10	0.019817	0.007622	0.007622	0.007622	0.007622	0.007622	0.007622	0.007622	0.007622	0.007622	0.007622	0.007622	0.007622	0.007622	0.007622	0.032012	0.007622	0.007622	0.007622
12		11	0.01625	0.00625	0.00625	0.01625	0.00625	0.00625	0.00625	0.02625	0.00625	0.00625	0.00625	0.00625	0.01625	0.00625	0.00625	0.00625	0.00625	0.00625	0.01625
13		12	0.005435	0.005435	0.005435	0.005435	0.222826	0.01413	0.022826	0.005435	0.005435	0.005435	0.005435	0.005435	0.040217	0.031522	0.005435	0.005435	0.005435	0.005435	0.144565
14		13	0.005787	0.005787	0.005787	0.015046	0.005787	0.005787	0.024306	0.005787	0.005787	0.005787	0.005787	0.015046	0.005787	0.005787	0.005787	0.015046	0.005787	0.005787	0.005787
15		14	0.005734	0.005734	0.005734	0.005734	0.014908	0.005734	0.005734	0.005734	0.005734	0.005734	0.005734	0.005734	0.005734	0.005734	0.005734	0.014908	0.005734	0.005734	0.244266
16		15	0.007911	0.007911	0.007911	0.007911	0.02057	0.007911	0.007911	0.007911	0.02057	0.007911	0.007911	0.007911	0.007911	0.007911	0.007911	0.007911	0.007911	0.007911	0.007911
17		16	0.008929	0.008929	0.008929	0.008929	0.008929	0.008929	0.008929	0.008929	0.008929	0.008929	0.008929	0.008929	0.008929	0.008929	0.008929	0.008929	0.008929	0.008929	0.008929
18		17	0.008446	0.008446	0.008446	0.021959	0.008446	0.008446	0.008446	0.008446	0.008446	0.008446	0.008446	0.008446	0.008446	0.008446	0.008446	0.008446	0.008446	0.008446	0.008446
19		18	0.009328	0.009328	0.009328	0.009328	0.009328	0.009328	0.009328	0.24254	0.009328	0.009328	0.009328	0.009328	0.009328	0.009328	0.009328	0.009328	0.009328	0.009328	0.009328
20		19	0.009058	0.009058	0.009058	0.009058	0.009058	0.009058	0.009058	0.009058	0.009058	0.009058	0.009058	0.009058	0.009058	0.009058	0.009058	0.009058	0.009058	0.009058	0.009058
21		20	0.007102	0.018466	0.007102	0.007102	0.007102	0.007102	0.007102	0.007102	0.007102	0.007102	0.018466	0.018466	0.007102	0.007102	0.007102	0.007102	0.007102	0.007102	0.007102
22		21	0.009058	0.009058	0.009058	0.009058	0.009058	0.009058	0.009058	0.009058	0.009058	0.009058	0.009058	0.023551	0.009058	0.009058	0.009058	0.009058	0.009058	0.009058	0.009058
23		22	0.008929	0.008929	0.008929	0.008929	0.008929	0.008929	0.008929	0.008929	0.008929	0.008929	0.008929	0.008929	0.008929	0.008929	0.008929	0.008929	0.008929	0.008929	0.008929
24		23	0.010246	0.010246	0.010246	0.010246	0.010246	0.010246	0.010246	0.010246	0.010246	0.010246	0.010246	0.010246	0.010246	0.010246	0.010246	0.010246	0.010246	0.010246	0.010246
25		24	0.005342	0.005342	0.005342	0.005342	0.022436	0.013889	0.005342	0.013889	0.005342	0.013889	0.005342	0.005342	0.167735	0.005342	0.005342	0.030983	0.005342	0.005342	0.005342
26		25	0.004921	0.004921	0.004921	0.004921	0.004921	0.004921	0.004921	0.036417	0.004921	0.004921	0.004921	0.004921	0.004921	0.004921	0.004921	0.004921	0.004921	0.004921	0.004921
27		26	0.00744	0.00744	0.00744	0.00744	0.00744	0.00744	0.00744	0.03125	0.00744	0.00744	0.00744	0.00744	0.019345	0.00744	0.00744	0.00744	0.00744	0.00744	0.00744
28		27	0.02983	0.007102	0.007102	0.007102	0.007102	0.007102	0.007102	0.052557	0.007102	0.007102	0.018466	0.018466	0.007102	0.007102	0.02983	0.007102	0.007102	0.007102	0.007102
29		28	0.007353	0.007353	0.007353	0.007353	0.007353	0.007353	0.007353	0.019118	0.007353	0.007353	0.007353	0.019118	0.007353	0.007353	0.007353	0.007353	0.007353	0.007353	0.007353
30		29	0.005081	0.013211	0.005081	0.005081	0.005081	0.005081	0.005081	0.005081	0.005081	0.005081	0.005081	0.013211	0.005081	0.005081	0.005081	0.005081	0.005081	0.005081	0.005081
31		30	0.006793	0.006793	0.006793	0.006793	0.006793	0.006793	0.006793	0.017663	0.006793	0.006793	0.017663	0.006793	0.006793	0.006793	0.006793	0.006793	0.006793	0.006793	0.006793
32		31	0.005952	0.005952	0.015476	0.005952	0.005952	0.015476	0.082143	0.005952	0.005952	0.005952	0.005952	0.034524	0.005952	0.005952	0.015476	0.005952	0.005952	0.005952	0.005952
33		32	0.121835	0.007911	0.007911	0.007911	0.007911	0.045886	0.02057	0.007911	0.007911	0.02057	0.007911	0.007911	0.007911	0.007911	0.007911	0.007911	0.007911	0.007911	0.007911
34		33	0.004845	0.004845	0.004845	0.004845	0.074612	0.004845	0.035853	0.004845	0.028101	0.004845	0.004845	0.012597	0.004845	0.004845	0.004845	0.020349	0.004845	0.004845	0.004845
35		34	0.007911	0.007911	0.007911	0.007911	0.02057	0.007911	0.007911	0.007911	0.007911	0.007911	0.007911	0.007911	0.007911	0.007911	0.007911	0.007911	0.007911	0.007911	0.007911

그림4.3에서 보이는 바와 같이 하나의 토픽은 모든 문서에 아주 작은 값이라도 가지고 있다. 이 결과를 그대로 사용한다면, 모든 토픽들은 모두 동

시 출현을 한다는 결론이 내려지기 때문에, 본 연구에서는 2015년 까지 우리나라에 등록된 특허중 화학 분야의 IPC코드를 가진 특허문서가 평균적으로 2.9개의 IPC코드를 가지고 있었으므로 한 문서의 상위 3개의 가중치 값을 가지는 토픽만을 유의미한 가중치를 가진다고 정의하였다. 그 결과 11,319건의 특허문서에서 존재 905,520번의 토픽들의 전체 출현 빈도수 중에서 51,492번의 토픽 출현 빈도수만이 유의미한 가중치를 가지게 되었다. 그림 4.4는 유의미한 가중치를 가지는 토픽은 1의 값을 할당 하고, 나머지는 0의 값을 할당 하여 만들어진 최종 문서-기술테마 매트릭스이며, X축은 3.1에서 정의한 바와 같이 LDA결과 토픽은 기술테마로 정의 되어 지므로, 토픽의 번호임과 동시에 외부기술테마의 번호이고, Y축은 문서의 번호이다. 이는 외부기술테마와 특허문서간의 출현 빈도수를 의미하며, ARM 수행에 있어서 기초 자료가 된다.

그림 4.4 문서-기술테마 매트릭스(일부)

#	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1																					
2	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	2	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0
4	3	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
5	4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	11	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
13	12	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
14	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
16	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	18	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
20	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
21	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	21	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
23	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
24	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
25	24	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0
26	25	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0
27	26	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	27	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
29	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
30	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
31	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
32	31	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
33	32	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
34	33	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
35	34	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

4.3.2 기술테마 간의 연관규칙 도출

본 과정은 외부기술테마들 간의 의미있는 규칙을 찾기 위해 ARM을 수행하는 과정이다. 본 연구에서는 총 문서에서 2%이상의 문서에서 조건부와 결론부의 기술테마가 동시 출현을 하여야 하고, 조건부가 출현한 문서 중 20% 이상의 문서에서 결론부의 기술테마가 함께 출현하여야 유의미한 규칙이라고 정의하기 위해 지지도와 신뢰도의 임계치를 각각 0.02, 0.2로 설정 하였으며, 향상도의 경우 양의 상관관계만을 고려하기 위해 1이상의 값을 가지는 규칙이어야 유의미한 규칙이라고 정의하였다. 그 결과 11,319건의 특허 데이터에서 유의미한 가중치를 가진 51,492개의 외부기술테마 중 96경우의 연관 규칙이 발굴 되었으며 발견된 규칙들은 부록3과 같으며, 표 4.4는 ARM결과의 일부 이다.

이때 지지도와 신뢰도의 임계치 값은 연관규칙을 100개 수준에서 뽑아 연구결과의 가시성을 높이기 위해 평균치 보다 조금 높은 값을 조정해 사용 하였으며, 임계치의 범위를 줄일수록 더 많은 규칙이 도출되어지지만, 약한 규칙이 함께 도출되어지고, 임계치의 범위를 늘일수록 더 적은 규칙이 도출되어지지만, 강한 규칙이 도출되어 진다.

표 4.4 ARM 결과(일부)

조건부		결론부	지지도	신뢰도	향상도
{topic01}	=>	{topic09}	0.0301	0.5430	5.2308
{topic02}	=>	{topic71}	0.0213	0.2513	5.7933
{topic02}	=>	{topic20}	0.0216	0.2555	4.4284
{topic02}	=>	{topic40}	0.0296	0.3493	4.4129
{topic02}	=>	{topic26}	0.0201	0.2377	2.7320
{topic04}	=>	{topic75}	0.0216	0.3664	4.5975

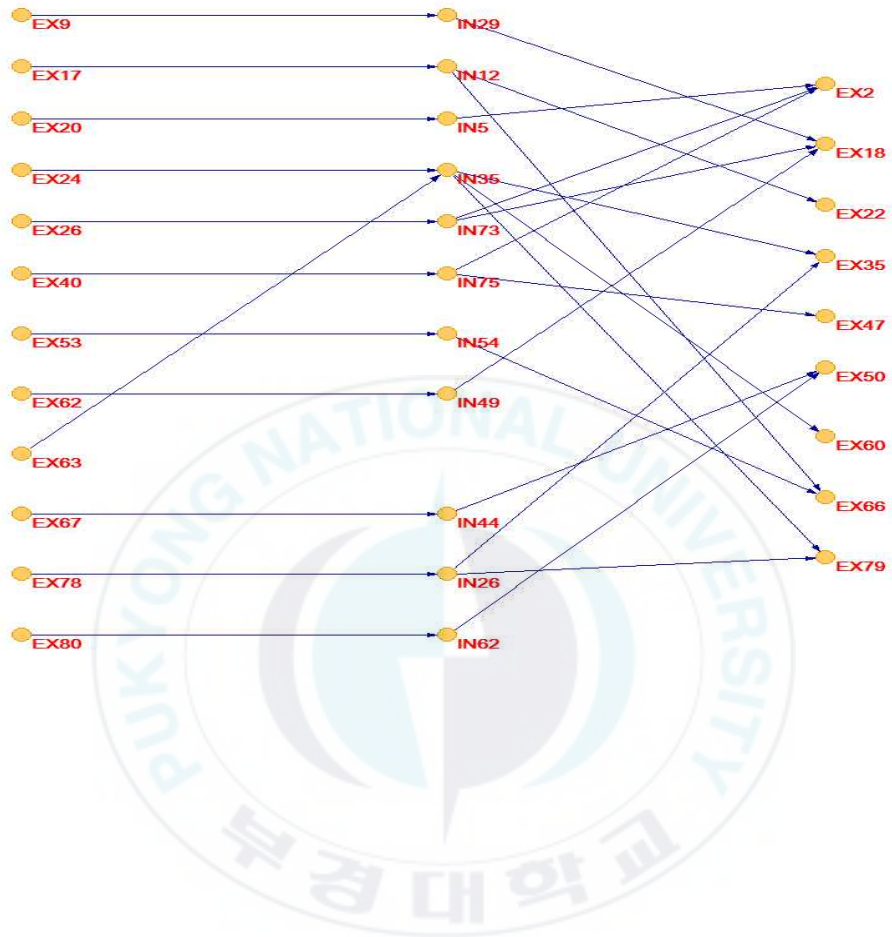
4.4 잠재적 기술테마 추천

4.4.1 최종 추천 네트워크

본 과정은 앞선 ARM결과에 유사도산출 결과를 반영하여 최종적인 추천 네트워크를 구성하는 과정으로, 앞선 96가지의 연관규칙 중 조건부에서 외부기술테마와 내부기술테마의 유사도가 낮을 경우 제외하고, 결론부에서 외부기술테마와 내부기술테마의 유사도가 높을 경우 제외함으로써, 내부기술을 기반으로 보유하지 않은 외부기술을 추천 받는 형태로 네트워크를 구성하고 있다.

ARM결과 조건부에 나타난 항목 중 Topic9, 17, 20, 24, 26, 40, 53, 62, 63, 67, 78, 80은 외부기술테마 이지만, 내부기술테마인 Topic29, 12, 5, 35, 73, 75, 54, 49, 35, 44, 26, 62와 유사도가 70% 이상으로 높기 때문에 유사도가 높은 외부기술테마 12개는 내부기술테마로써 간주 되어 질 수 있다. 이후 유사도가 높은 12개의 기술테마를 제외한 나머지 항목들을 제거하고, 기존의 96경우의 연관 규칙 중 남은 35경우의 연관 규칙에서, 결론부에 나타난 항목 중 Topic2, 18, 22, 35, 47, 50, 60, 66, 79의 외부기술테마는 내부기술테마와 유사도가 70% 이하로 낮기 때문에 유사도가 낮은 9개의 기술테마를 제외한 나머지 항목들은 제외 하는 작업을 거쳐 35경우의 연관 규칙 중 17경우의 연관 규칙에 의해 9가지의 기술테마가 최종적인 잠재 추천 기술로써 도출되어 졌다. 그림4.5는 최종적인 추천 네트워크의 결과를 도식화 한 것이다.

그림 4.5 최종 추천 네트워크



4.4.2 결과 해석

앞선 결과에서 최종적인 추천 네트워크가 만들어 졌으며, 17경우의 연관 규칙에 의해 최종적으로 9가지의 외부기술테마가 추천 후보 군으로 도출되었다. 하지만, 지금까지 과정에서 기술테마들은 LDA의 결과인 토픽을 구성하고 있는 10개의 단어 형태로 존재하였다. 따라서 본 과정에서 단어들의 의미와 연관관계를 통해 각 테마들의 명칭을 정하는 작업을 진행 한 후, 결과 해석을 진행한다.

먼저 조건부에 해당하는 기술테마의 명칭을 정하는 작업을 진행한다. 이때 ARM의 결과에서 조건부에 해당하는 기술테마 들은 외부기술테마 이기 때문에 유사도가 높은 내부기술테마의 단어들로 토픽의 명칭을 정한다. 표 4.5는 외부기술테마와 내부기술테마의 비교를 위한 예시이다. 앞선 4.2의 유사도계산에서 외부기술테마 9번과 내부기술테마 29번의 유사도가 82%였으며 내부기술테마와 외부기술테마 단어의 구성을 보면 매우 유사하다는 것을 알 수 있다. 그렇기 때문에 보다 명확한 결과해석을 위하여 조건부에 해당하는 외부기술테마는 유사도가 높은 내부기술테마의 명칭으로 대체한다. 이와 같은 과정을 통해 정해진 내부기술테마의 명칭은 표4.6과 같고, 외부기술테마의 명칭은 표4.7과 같다.

표 4.5 내부기술테마와 외부기술테마의 단어 비교

내부기술 테마29	crystal	liquid	display	alignment	film
	composition	optical	angle	color	viewing
외부기술 테마9	liquid	crystal	display	alignment	film
	substrate	substrates	layer	pair	panel

표 4.6 내부기술테마 명칭

2차 전지 분리 막 기술				
electrolyte	secondary	battery	lithium	nonaqueous
solvent	salt	solution	characteristics	additive
빛 입자				
optical	light	filter	element	image
color	display	stereoscopic	mask	manufacturing
화합물				
compound	comprising	aliphatic	nitrile	residual
mixture	nitrogen	forms	polym	solubility
광학				
crystal	liquid	display	alignment	film
composition	optical	angle	color	viewing
화합물 반도체				
group	formula	compound	represented	consisting
chemical	disclosed	semiconductor	alkyl	comprising
고분자 화합물				
weight	parts	molecular	based	comprising
average	ratio	mixture	acid	total
편광판				
film	polarizer	optical	thin	angle
retardation	linear	films	lcd	integrated
발광체				
organic	light	emitting	electronic	hole
derivative	efficiency	electron	injecting	excellent
열경화성수지				
composition	excellent	resistance	curable	surface
thermal	curing	hardness	exhibit	property
삼투압				
layer	layers	forming	lay	sequentially
diffusion	element	therebetween	waterrepelling	disclosed
2차 전지 음극재				
lithium	oxide	active	metal	transition
cathode	capacity	mixed	composite	higher

표 4.7 외부기술테마 명칭

비 수성 리튬 2차전지 전해질				
electrode	active	lithium	battery	secondary
positive	negative	electrolyte	collector	nonaqueous
디스플레이				
electrode	pixel	line	insulating	gate
transistor	electrodes	common	substrate	lines
발광체				
radiation	electromagnetic	sample	arrangement	frequency
based	generate	multiple	time	characteristic
제약, 의약				
compounds	compositions	formula	relates	methods
pharmaceutical	treating	salts	defined	conditions
올레핀 촉매				
catalyst	relates	preparing	complex	olefin
ligand	reaction	preparation	ethylene	comprising
감광성 에폭시수지				
resin	composition	compound	photosensitive	specific
epoxy	thermoplastic	curing	obtained	molded
제약, 의약				
compounds	treatment	relates	methods	cancer
compositions	activity	comprising	diseases	pharmaceutical
백라이트				
light	source	emitting	wavelength	reflective
backlight	color	led	sources	reflected
제약, 의약				
salt	acceptable	disease	pharmaceutically	diseases
treatment	treating	effective	preventing	agent

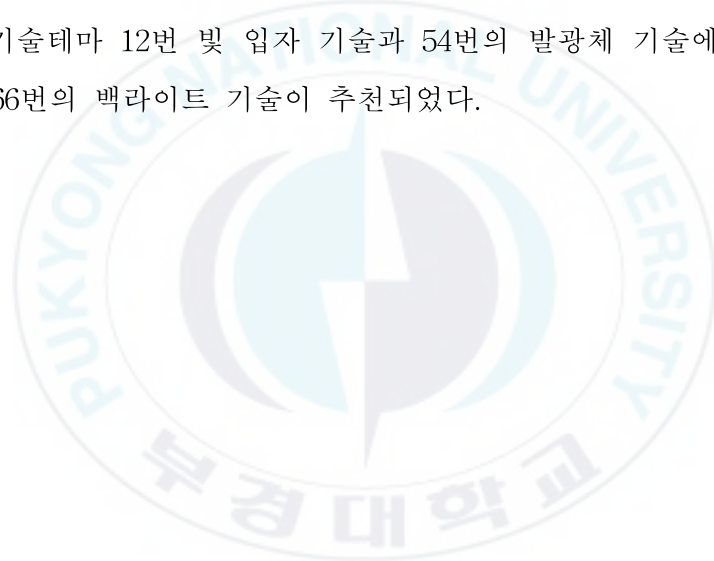
본 연구의 최종결과 내부기술테마 5번의 2차 전지 분리 막 기술과 73번 삼투압 기술, 75번의 2차 전지 음극재 기술에 의해 외부기술테마 2번의 비수성 리튬 2차 전지 전해질 기술이 추천 되었다. 이는 최근 이슈가 되고 있는 액체 전해질을 사용한 배터리 폭파 사고에서 알 수 있듯이, 많은 배터리 제조업체들은 액체 전해질의 위험성을 인식하고 비수성 전해질 개발에 힘쓰고 있어 이러한 결과가 나온 것이라고 고려되어진다.

두 번째로 내부기술테마 29번의 광학 기술과 73번의 삼투압 기술, 49번의 편광판 기술에 의해 외부기술테마 18번의 디스플레이에 관한 기술이 추천 되었다. 이는 LG화학이 화학 회사이기 때문에 기본적인 디스플레이제품과 관련된 광학기술과 이와 관련된 기초적인 화학 기술 등은 보유하고 있지만, 실제로 제품을 만들어 파는 등의 사업화 활동은 하지 않기 때문에 해당 특허를 보유하지 않아 나온 결과라 고려되어지며, 추후 사업영역 확장이나 M&A에 있어서 의사결정시 고려할 가치가 있다고 생각된다.

세 번째로 내부기술테마 35번의 화합물 반도체 기술과 26번의 화합물기술에 의해 외부기술테마 35번, 60번, 79번의 제약, 의약 기술이 추천 되었다. 앞서 기술테마의 명칭을 정하는 과정에서 35번, 60번, 79번 모두 제약, 의약으로 명칭을 정해였으나, 3가지의 기술테마의 단어들은 공통적으로 제약과 관련된 단어들이임에도 불구하고 약간의 차이를 가지고 있다. 35번의 경우 compounds, compositions와 같이 제약과 직접적으로 관련된 기술으로 보이며, 60번의 경우 treatment, cancer과 같이 직접적인 치료제와 관련된 기술로 보인다. 또한 79번의 경우 effective, preventing, agent와 같은 단어를 보아 질병을 예방하기 위한 약품을 만드는 기술로 고려되어진다. 이와 같이 LG화학의 화학물질을 합성하는 등의 기술을 토대로 제약과 관

련된 사업에도 확장이 가능할 것으로 보이며, 실제 LG화학과 관련된 공시 등에 의해 LG생명과학과 최근 인수합병을 준비 중이라는 사실을 알 수 있었다.

또한 내부기술테마 12번 빛 입자 기술에 의해 외부기술테마 22번 발광체 기술, 내부기술테마 75번의 2차 전지 음극재 기술에 의해 외부기술테마 47번의 올레핀 촉매 기술, 내부기술테마 44번 고분자 화합물기술과 62번의 열경화성 수지 기술에 의해 외부기술테마 50번의 감광성 에폭시 수지 기술, 내부기술테마 12번 빛 입자 기술과 54번의 발광체 기술에 의해 외부 기술테마 66번의 백라이트 기술이 추천되었다.



4.4.3 상세 특허 추천

최종적으로 추천된 외부기술테마는 2, 18, 22, 35, 47, 50, 60, 66, 79번으로써, 해당하는 기술테마의 가중치가 높은 상위 3개의 특허 리스트를 제시한다. 표4.8은 최종 추천 특허 리스트이며, 발명의 명칭은 같으나 출원 번호가 다른 경우 하나의 특허만 명시 하였다.

표 4.8 추천 특허 리스트

비 수성 리튬 2차 전지 전해질	
별명의 명칭	출원 번호
Cable-type secondary battery	14/483318
Non-aqueous battery with columnar active material	12/064307
Lithium-ion battery	10/979043

디스플레이	
별명의 명칭	출원 번호
Liquid crystal display device	13/449871
In-plane switching mode liquid crystal display device	11/166922
Thin film transistor array panel for liquid crystal display having pixel electrode	11/873767

발광체	
별명의 명칭	출원 번호
Radiance processing by demultiplexing in the frequency domain	12/186392
Method and apparatus for performing optical imaging using frequency-domain interferometry	12/795529

제약, 의약	
별명의 명칭	출원 번호
Processes for preparing 1,2-substituted cyclopropyl derivatives	13/232751
Compounds as cannabinoid receptor ligands and uses thereof	12/274105
Oxopiperazine-azetidine amides and oxodiazepine-azetidine amides as monoacylglycerol lipase inhibitors	13/246232
Glycoside derivatives and uses thereof	13/425888

올레핀 촉매	
별명의 명칭	출원 번호
Process for generating linear alpha olefin comonomers	11/517871
Oxide catalyst and phosphoric oxide catalyst for hydrocarbon steam cracking, method for preparing the same and method for preparing olefin by using the same	11/902665

감광성 에폭시수지	
별명의 명칭	출원 번호
Epoxy resin composition for sealing, and electronic component device	13/976538
Cationic electrodeposition coating composition	13/369350
Flame resistant polyester compositions, method of manufacture, and articles thereof	12/826013

제약, 의약	
별명의 명칭	출원 번호
1',3'-disubstituted-4-phenyl-3,4,5,6-tetrahydro-2H,1'-H-[1,4'-bipyridinyl-2'-ones	13/803267
1,3-disubstituted-4-phenyl-1H-pyridin-2-ones	12/677628
Di(arylamino)aryl compounds	12/448759

백라이트	
별명의 명칭	출원 번호
Display illumination device with a film-based lightguide having stacked incident surfaces	13/088167
Light source device, imaging apparatus and endoscope apparatus	12/484198
Backlight device	11/667397

제약, 의약	
별명의 명칭	출원 번호
Nitrogenous heterocyclic derivative and medicine containing the same as an active ingredient	13/682784
7-piperidinoalkyl-3, 4-Dihydroquinolone derivative	13/122281
Nitrogenous heterocyclic derivative and medicine containing the same as an active ingredient	13/016849

5. 결 론

본 연구는 기업의 내부역량을 바탕으로 새로운 기술기회 발굴을 위한 방법론을 제시하는 연구로써, 기존의 기술기회 발굴 연구가 대부분 내부역량을 고려하지 않은 채 연구를 진행하였거나 기술기회 발굴의 결과로써 IPC와 같이 세부 기술 단위의 분야가 아닌 대분류 단위의 특허분류코드 등을 제시하기 때문에 공공기관이나 민간기업의 연구개발, 기술사업화 등의 의사결정에 있어서 상세기술을 알기 어려워 실제 활용하는 것이 어렵다는 한계점 지니고 있었다. 그렇기 때문에 본 연구에서는 이를 해결하고자 토픽모델링의 LDA와 ARM 기법을 적용하여 기업의 내부역량을 고려한 후, 특허가 포함하고 있는 기술을 기술테마의 단위로 세분화 한 후 추천 네트워크를 구성함으로써, 실제 활용이 용이한 결과를 도출했다는 점에 의의가 있다.

또한 최근 급변하는 시장 환경 속에서 글로벌 기술경쟁의 우위를 점하기 위해 새로운 기술기회를 발굴하고 선점하는 것이 중시됨에 따라, 실현 가능한 기술기회를 발굴하기 위한 방법론을 개발하는 것은 기업의 지속적 성장을 위한 필수요소가 되었다. 이에 본 연구에서 제안 하고 있는 방법론은 보유 특허를 기반으로 내부역량을 고려하고, ‘기술테마’라는 비교적 상세한 단위로 새로운 기술기회를 제안함과 더불어 상세특허 리스트를 제시하기 때문에 공공연구기관의 기술가치 평가, 기술사업화 등의 의사결정과 민간기업의 사업영역 확장이나 M&A 등의 의사결정에 있어서 많은 시간과 비용을 절약할 수 있을 것 이라 기대되어진다.

하지만, 이와 같은 연구 의의에도 불구하고 본 연구는 한계점을 지니고 있으며, 추후 연구를 진행함으로써 한계점을 극복 하고자 한다.

첫째, 본 연구에서 LDA를 수행함에 있어 코퍼스를 구성할 때, 오직 등록된 특허의 요약문만을 이용하였다. 특허는 출원 에서 공개까지 약 1년 6개월의 기간이 소요되고, 등록까지는 더 많은 시간이 소요되어질 수 있다. 그렇기 때문에 본 연구 방법론의 결과를 통해 추천 받은 기술테마 또한 과거의 기술력일 수 있기 때문에 급변하는 시장 속 에서 뛰어난 경쟁력을 지닌 기술을 추천받는 것은 다소 어려울 수 있다.

둘째, 본 연구의 방법론은 등록 특허의 기술테마 추출을 한 후 연관규칙을 발견하여 결과를 제시하고 있다. 본 연구의 ARM과정에 있어 조건부는 내부 역량에 기반 한 기술테마, 결론부는 내부 역량을 바탕으로 검색된 타 기업의 특허의 기술테마들로 구성되어 있다. 그렇기 때문에 추천 받은 기술테마 역시 타 기업 특허 문서에서 발췌된 기술테마 일 것이며, 이는 이미 해당 기술 분야는 이미 많은 기업들의 경쟁이 심화되어 있어 진입 장벽이 매우 높을 수 있다는 것을 의미 할 가능성도 있다.

셋째, 연구의 과정에 있어 기술테마의 개수를 임의적으로 80개로 지정하였기 때문에, 80개의 테마가 코퍼스를 구성하고 있는 모든 문서의 핵심 내용을 포함하지 못하고 있을 수 있거나 반대로 코퍼스를 구성하고 있는 문서의 수에 대비하여 너무 많은 기술테마가 추출되어 중복된 기술테마가 존재할 수 있다.

마지막으로, 기술테마의 명칭을 정하는 과정에 있어서 단어의 구성에 의존하여 주관적인 판단으로 정하였기 때문에, 본 연구의 프로세스를 진행하는 사람에 따라 기술테마의 명칭이 달라질 수 있다. 이는 연구의 객관성과 신뢰성을 줄일 수 있음을 의미할 수 있다.

이와 같은 한계점들을 극복하기 위해 추후 연구에서는 코퍼스를 구성할 때 등록특허 뿐만 아니라 해당 분야의 과학, 기술의 정보를 포함한 논문 혹은 저널 등을 포함 하여 추천된 기술의 다양성을 좀 더 높일 수 있는 방안을 고려 할 것이며, LDA를 활용한 연구 인 연도별 토픽트렌드 연구를 접목하여 추천된 기술 경쟁 심화 정도를 예측해 볼 것이다. 또한 적절한 기술테마의 개수를 산정하기 위하여, LDA를 수행할 때 적절한 토픽의 수를 정하기 위해 Griffiths와 Steyvers가 제안한 Finding scientific topics연구를 접목하여 합리적인 토픽의 수를 정할 것이며, 기술테마의 명칭을 정하기 위하여 구글의 연구원인 Tomas Mikolov와 Kai Chen, Greg Corrado, Jeffrey Dean에 의해 제안된 word2vec기법을 적용함으로써, 자동화된 프로세스를 가능하게 하여 연구의 객관성과 신뢰성을 높이기 위해 노력할 것이다.

참고문헌

- [1] 김무진 & 윤장혁. (2014). 핵심특허 발전도 생성 방법: Latent Dirichlet Allocation 과 인용정보의 활용. 2014 년 대한산업공학회 추계학술대회 논문집, 1653-1668.
- [2] 김정석 & 이영덕. (2009). 기술예측을 통한 미래 유망기술 우선순위 평가모형에 관한 연구. 한국기술혁신학회 2009 년 추계학술대회, 109-127.
- [3] 김창석 & 신준석. (2013). 기업 기술역량 기반 기술기회탐색에 관한 연구. 2013 년 대한산업공학회 추계학술대회 논문집, 1249-1258.
- [4] 노대민 & 김종주. (2013). 기술수명주기와 제품수명주기의 비교 시장구조와 기업의 연구개발 특성을 중심으로. 서비스마케팅저널, 6(1), 21-41.
- [5] 노현정 & 임효정. (2009). 특허정보를 활용한 나노바이오 융합기술의 특성 분석. 지식재산연구, 4(3), 109-129.
- [6] 박영욱, 홍성화, 이준영, 여운동 & 권오진. (2011). 미래유망기술에 대한 정량적 평가방법 연구. 한국기술혁신학회 2011 년 추계학술대회, 3-12.

- [7] 박영진, 고남욱 & 윤장혁. (2015). 보유특허 기반의 기술기회탐색을 위한 특허추천방법. 지식재산연구, 10(1), 169-200.
- [8] 박현석, 서원철, 고병열, 이재민 & 윤장혁. (2014). 기업의 보유 기술 및 제품에 기반한 기술기회발굴. 대한산업공학회지, 40(5), 442-450.
- [9] 서원철, 윤장혁, 고병열, 이재민 & 송인석. (2014). 보유 제품 포트폴리오 기반의 제품기회발굴 방법론. 2014 년 대한산업공학회 추계 학술대회 논문집, 2596-2601.
- [10] 윤병운, 박인채, 이호신 & 고병열, (2011). 기술 인텔리전스를 활용한 기술기회탐색: 유형화, 현황, 통합체계개발을 중심으로. 한국기술혁신학회지, 14, 1073-1095.
- [11] 이세준, 이윤준 & 홍정임. (2008). 통합적 미래연구 방법론의 탐색 및 적용. 정책연구, 1-229.
- [12] 정병기 & 윤장혁. "토픽모델링과 감성분석을 활용한 제품 개발 기회 도출." 대한산업공학회 춘계공동학술대회 논문집 (2016): 724-728.
- [13] 정병기, 김정욱 & 윤장혁. "토픽모델링을 활용한 증강현실기술의 특허경쟁정보분석." 대한산업공학회 추계학술대회 논문집 (2015): 2265-2270.
- [14] 조은경 & 한영균. (2014). 텍스트 말뭉치의 주제와 주제어 분석-세

중 문어 말뭉치를 중심으로. 텍스트언어학, 37(단일호), 201-232.

- [15] 최하영, 고남욱 & 윤장혁. "연관규칙분석을 활용한 융합 기술 기회 발굴 방법론." 대한산업공학회 추계학술대회 논문집 (2015): 870-874.
- [16] 최하영 & 윤장혁. (2016). 특허의 발명 컨셉 마이닝을 통한 기술 기회 분석. 대한산업공학회 춘계공동학술대회 논문집, 129-132.
- [17] Abernathy, W. J. & Clark, K. B. (1985). Innovation: Mapping the winds of creative destruction. Research policy, 14(1), 3-22.
- [18] Agrawal, R. Imieliński, T. & Swami, A. (1993, June). Mining association rules between sets of items in large databases. In Acm sigmod record, 22(2), 207-216.
- [19] Agrawal, R.. & Srikant, R. (1994, September). Fast algorithms for mining association rules. In Proc. 20th int. conf. very large data bases, 1215, 487-499.
- [20] Kim, C., Lee, H., Seol, H. & Lee, C. (2011). Identifying core technologies based on technological cross-impacts: An association rule mining (ARM) and analytic network process (ANP) approach. Expert Systems with Applications, 38(10), 12559-12564.
- [21] Choi, S., Yoon, J., Kim, K., Lee, J. Y. & Kim, C. H. (2011).

SAO network analysis of patents for technology trends identification: a case study of polymer electrolyte membrane technology in proton exchange membrane fuel cells. *Scientometrics*, 88(3), 863–883.

- [22] Christensen, C. (2013). *The innovator's dilemma: when new technologies cause great firms to fail*. Harvard Business Review Press.
- [23] Blei, D. M., Ng, A. Y. & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan), 993–1022.
- [24] Klevorick, A. K., Levin, R. C., Nelson, R. R. & Winter, S. G. (1995). On the sources and significance of interindustry differences in technological opportunities. *Research policy*, 24(2), 185–205.
- [25] Krestel, R., Fankhauser, P. & Nejdl, W. (2009). Latent dirichlet allocation for tag recommendation. In *Proceedings of the third ACM conference on Recommender systems*, 61–68.
- [26] Lee, M., Lee, S., Lee, J., Jung, H., Kim, J., Seo, D. & Sung, W.K. (2013). InSciTe Advanced: Strategic Decision-Making Support Service based on Technology Opportunity Discovery Model. *INFORMATION-AN INTERNATIONAL INTERDISCIPLINARY JOURNAL*, 16(1 B), 693–698.
- [27] Shmueli, G., Patel, N. R. & Bruce, P. C. (2007). *Data mining for*

business intelligence: concepts, techniques, and applications in Microsoft Office Excel with XLMiner. John Wiley & Sons.

- [28] Steyvers, M. & Griffiths, T. (2007). Probabilistic topic models. *Handbook of latent semantic analysis*, 427(7), 424-440.
- [29] Tang, J., Wang, B., Yang, Y., Hu, P., Zhao, Y., Yan, X.. & Usadi, A. K. (2012, August). PatentMiner: topic-driven patent analysis and mining. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 1366-1374.
- [30] Yoon, J. & Kim, K. (2011). Detecting signals of new technological opportunities using semantic patent analysis and outlier detection. *Scientometrics*, 90(2), 445-461.
- [31] Yoon, J, Kim, K(2011), Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks, *Scientometrics*, 88(1), 213-228.
- [32] Zha, X. & Chen, M. (2010). Study on early warning of competitive technical intelligence based on the patent map. *Journal of Computers*, 5(2), 274-281.
- [33] Zhu, D. & Porter, A. L. (2002). Automated extraction and visualization of information for technological intelligence and forecasting. *Technological*

forecasting and social change, 69(5), 495-506.



부록

1. 내부기술테마

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
materials	battery	high	monomer	electrolyte
application	cells	due	resin	secondary
thermoelectric	module	conductivity	copolymer	battery
conversion	modules	mechanical	vinyl	lithium
plastic	mounted	ionic	latex	nonaqueous
devices	pack	simple	impact	solvent
films	plurality	strength	rubber	salt
substrates	side	easy	aromatic	solution
optoelectronic	case	low	thermoplastic	characteristics
substrate	members	solid	graft	additive

Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
column	acid	electrochemical	water	carbon
distillation	methacrylic	stability	phosphate	core
wall	step	thermal	iron	component
conventional	solution	mixture	solution	crystalline
dividing	gas	separation	controller	coating
point	obtained	improving	dyeing	low
compared	alcohol	eutectic	iodine	consisting
zone	aqueous	center	tim	amorphous
highpurity	based	excellent	additives	capable
energy	continuous	amide	carbon	silicon

Topic 11	Topic 12	Topic 13	Topic 14	Topic 15
resin	optical	properties	agent	temperature
composition	light	prepared	flame	high
photosensitive	filter	low	retardant	heating
resistance	element	compounds	styrene	discharging
relates	image	superior	groups	charging
excellent	color	physical	polycarbonate	life
binder	display	ester	makes	room
good	stereoscopic	therefrom	react	long
mechanical	mask	insulating	rigidity	polyurethane
polyimide	manufacturing	silane	basic	highpower

Topic 16	Topic 17	Topic 18	Topic 19	Topic 20
preparing	surface	manufacturing	type	electric
capable	area	reduced	time	disclosed
lactate	transparent	manufactured	driving	insulation
gene	opening	article	mandrel	hybrid
pla	ratio	improve	controlled	pouch
copolymer	electroconductive	time	exhibits	vehicles
microorganism	electrical	productivity	addition	foreign
introduced	entire	blend	disclosed	andor
lactylcoa	conductor	multilayered	efficiently	contact
mutant	work	photovoltaic	increasing	source

Topic 21	Topic 22	Topic 23	Topic 24	Topic 25
battery	portion	conductive	negative	member
case	sealing	energy	retardation	assembly
part	portions	level	positive	portion
secondary	tape	ntype	index	disposed
contact	attached	ptype	relates	coupled
receiving	pouch	comprises	refractive	extending
opposite	side	difference	thickness	aperture
outer	guide	interposed	compensation	wall
mounted	welded	relates	characteristic	plurality
contacting	machine	disposed	wavelength	tubular

Topic 26	Topic 27	Topic 28	Topic 29	Topic 30
compound	bus	lower	crystal	battery
comprising	bar	upper	liquid	connection
aliphatic	wire	frame	display	member
nitrile	bars	member	alignment	terminal
residual	laser	parts	film	cell
mixture	connector	horizontal	composition	connected
nitrogen	coupled	surfaces	optical	circuit
forms	fluorine	rate	angle	electrically
polym	minimized	vertical	color	terminals
solubility	locking	ethylene	viewing	insulative

Topic 31	Topic 32	Topic 33	Topic 34	Topic 35
membrane	safety	disclosed	reaction	group
electrolyte	cap	resistance	reactor	formula
fuel	assembly	isolation	unsaturated	compound
ion	top	fault	producing	represented
reverse	gasket	function	mixture	consisting
conductivity	bent	ground	aldehydes	chemical
amine	element	specific	efficiency	disclosed
moietycontaining	vent	selfdiagnosing	separating	semiconductor
osmosis	cylindrical	diagnosis	performing	alkyl
polyamide	pressure	improving	production	comprising

Topic 36	Topic 37	Topic 38	Topic 39	Topic 40
adhesive	coating	unit	disclosed	battery
excellent	porous	units	molding	pack
acrylic	particles	plurality	capable	total
property	binder	number	mold	slave
relates	inorganic	backlight	effectively	management
composition	substrate	arranged	injection	bms
pressure	separator	increased	comprising	master
durability	layer	contact	conditions	communication
high	surface	condensing	insert	system
pressuresensitive	polymer	input	overdischarge	target

Topic 41	Topic 42	Topic 43	Topic 44	Topic 45
specific	catalyst	gas	weight	battery
panel	olefin	potential	parts	cooling
artificial	preparing	comprises	molecular	module
surface	relates	generation	based	heat
composite	high	disclosed	comprising	channel
marble	hydrocarbon	charging	average	housing
gravity	cracking	generated	ratio	fluid
low	polymerization	improving	mixture	disposed
chips	metallocene	caused	acid	flow
high	steam	plateau	total	fin

Topic 46	Topic 47	Topic 48	Topic 49	Topic 50
electrode	amount	plate	film	pattern
active	groups	base	polarizer	substrate
secondary	disclosed	polarizing	optical	ink
layer	viscosity	plates	thin	line
outer	trihalosilane	side	angle	forming
collector	due	protective	retardation	printing
surface	controlled	disposed	linear	patterns
cabletype	ink	thickness	films	transparent
battery	small	vertically	lcd	width
polarity	multifunctional	films	integrated	head

Topic 51	Topic 52	Topic 53	Topic 54	Topic 55
preparing	electrode	metal	organic	body
toner	assembly	transition	light	main
paper	leads	complex	emitting	housing
polymerized	lead	group	electronic	cover
improved	electrodes	nanoparticles	hole	edge
ceramic	tabs	bridge	derivative	axis
excellent	constructed	density	efficiency	parallel
characteristics	separator	large	electron	distance
charge	connected	ligand	injecting	ends
efficiency	disposed	amido	excellent	electrical

Topic 56	Topic 57	Topic 58	Topic 59	Topic 60
power	glass	part	battery	composition
charge	molten	fixed	voltage	addition
control	bath	housing	estimating	monomer
capacity	float	cone	resistance	norbornene
signal	capable	spinning	soc	solvent
operating	bottom	axis	data	producing
controller	metal	condition	calculating	produced
line	casing	rotation	charge	reaction
operation	tin	side	measured	content
supply	etching	supply	measuring	polymer

Topic 61	Topic 62	Topic 63	Topic 64	Topic 65
battery	composition	voltage	cathode	sheet
pack	excellent	battery	anode	disclosed
air	resistance	discharge	active	sheets
vehicle	curable	sensing	secondary	laminate
external	surface	unit	separator	size
control	thermal	charging	coated	base
medium	curing	connected	mix	cutting
controlling	hardness	measuring	coating	step
range	exhibit	switching	activematerial	continuous
order	property	terminal	collector	direction

Topic 66	Topic 67	Topic 68	Topic 69	Topic 70
slurry	copolymer	coolant	polymer	heat
polishing	block	port	polymerization	tube
cmp	relates	inlet	monomer	exchanger
disclosed	olefin	space	photoreactive	exchange
silicon	repeating	flow	preparing	dissipation
abrasive	based	outlet	gel	transfer
adjuvant	preparation	direction	initiator	inside
charged	properties	discharged	rate	performance
increase	distribution	rotary	relates	generated
selectivity	molecule	super	crosslinking	medium

Topic 71	Topic 72	Topic 73	Topic 74	Topic 75
comprising	roll	layer	comprises	lithium
disclosed	board	layers	comprising	oxide
andor	excellent	forming	relates	active
reduction	metallic	lay	alkali	metal
shows	paste	sequentially	terephthalaldehyde	transition
combination	photochromic	diffusion	oxides	cathode
totally	printed	element	significantly	capacity
hollow	hard	therebetween	alkalifree	mixed
groupcontaining	spray	waterrepelling	exhibiting	composite
component	paint	disclosed	residue	higher

Topic 76	Topic 77	Topic 78	Topic 79	Topic 80
battery	cell	voltage	cerium	particle
system	battery	circuit	powder	particles
time	balancing	microprocessor	oxide	diameter
vector	plurality	level	preparing	range
plurality	solar	signal	precursor	size
estimated	stack	driver	solution	average
indicative	adjacent	generating	carbonate	copper
determining	abnormality	diagnostic	solvent	reducing
manufacture	alignmentcoupling	difference	mixing	dispersion
based	assembly	electrical	prepared	excellent



2. 외부기술테마

표 4.3 외부기술테마

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
direction	electrode	block	portion	pressure
angle	active	copolymer	portions	flexible
axis	lithium	membrane	aperture	input
polarizing	battery	oxygen	disposed	techniques
polarizer	secondary	comprising	edge	computing
optical	positive	segment	receiving	higher
index	negative	improved	extending	hinge
parallel	electrolyte	hydrophobic	face	sensitive
polarization	collector	copolymers	holder	respective
retardation	nonaqueous	polyethylene	center	hydrocarbons

Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
comprising	high	surface	liquid	element
comprises	temperature	surfaces	crystal	elements
free	low	length	display	capable
polyol	ratio	prevent	alignment	nanotube
foam	range	bonding	film	metallic
polyurethane	density	surfac	substrate	nanotubes
species	rate	major	substrates	deformable
isocyanate	characteristics	underlying	layer	prevents
polyether	stability	exposed	pair	fabric
prepolymer	large	treated	panel	bit

Topic 11	Topic 12	Topic 13	Topic 14	Topic 15
conversion	group	semiconductor	plurality	unit
reduce	carbon	laser	characteristics	units
solar	atoms	capable	barrier	number
transmission	groups	aspect	arranged	structural
conventional	compound	peak	adjacent	repeating
photovoltaic	cyclic	probe	imager	derived
column	consisting	providing	imagers	stage
photoelectric	bonded	nanocrystal	generated	relationship
maximum	chain	pulse	enhanced	roll
achieve	linear	diffraction	images	small

Topic 16	Topic 17	Topic 18	Topic 19	Topic 20
battery	optical	electrode	particles	solution
cell	image	pixel	particle	solvent
cells	filter	line	size	solid
module	color	insulating	diameter	salt
case	light	gate	range	electrolyte
pack	array	transistor	average	ion
modules	lens	electrodes	primary	fuel
mounted	imaging	common	shell	additive
disclosed	elements	substrate	distribution	electrolytic
sensing	beam	lines	fine	carbonate

Topic 21	Topic 22	Topic 23	Topic 24	Topic 25
pattern	radiation	disclosed	compound	ink
resist	electromagnetic	application	formula	medium
component	sample	specifically	represented	printing
composition	arrangement	die	general	recording
solubility	frequency	disclos	consisting	pigment
generator	based	providing	group	toner
alkali	generate	primary	defined	inkjet
photoresist	multiple	manufacture	chemical	dye
action	time	holographic	formulae	media
capable	characteristic	thermoelectric	iii	printed

Topic 26	Topic 27	Topic 28	Topic 29	Topic 30
layer	fluid	conductive	structures	body
layers	flow	transparent	components	main
hole	chamber	protective	methods	cover
injection	air	electrically	making	section
contact	cooling	plastic	transfer	wall
transistors	valve	article	features	opening
lay	port	screen	compositions	container
transport	inlet	diodes	embodiments	housing
dielectric	heat	thermally	feature	space
thickness	outlet	conducting	applications	attached

Topic 31	Topic 32	Topic 33	Topic 34	Topic 35
embodiments	forming	contact	sensor	compounds
control	step	electrical	data	compositions
values	manufacturing	terminal	analyte	formula
environment	exposure	connection	unit	relates
pump	mask	connected	calibration	methods
methods	steps	electrically	control	pharmaceutical
additional	production	connector	disposed	treating
volume	etching	board	based	salts
order	manner	terminals	reference	defined
disclosed	cost	connecting	housing	conditions

Topic 36	Topic 37	Topic 38	Topic 39	Topic 40
methods	acid	frame	member	metal
systems	pathway	front	assembly	oxide
devices	amount	member	outer	transition
host	period	vehicle	coupled	composite
measurement	microbial	members	disposed	earth
measuring	enzyme	mounting	coupling	metals
relates	time	rear	housing	mixed
glucose	nucleic	assembly	shaft	complex
embodiments	encoding	pair	sealing	consisting
sensing	organism	coupled	central	titanium

Topic 41	Topic 42	Topic 43	Topic 44	Topic 45
data	charge	electrodes	agent	and/or
system	carrier	lead	coating	clutch
communication	magnetic	system	composition	relative
network	field	medical	water	orientation
control	comprises	patient	dispersion	combined
input	photochromic	wire	aqueous	amount
computer	coil	delivery	formation	accessory
wireless	cylinder	leads	coated	selective
output	activated	implantable	surfactant	object
processing	relates	connector	silane	variable

Topic 46	Topic 47	Topic 48	Topic 49	Topic 50
area	catalyst	point	base	resin
energy	relates	single	oil	composition
storage	preparing	relates	composition	compound
condition	complex	generation	lubricating	photosensitive
means	olefin	greater	silicone	specific
areas	ligand	comprising	viscosity	epoxy
defined	reaction	roof	glycol	thermoplastic
providing	preparation	mol	lubricant	curing
prevent	ethylene	interface	engine	obtained
stored	comprising	ethylene	phosphorus	molded

Topic 51	Topic 52	Topic 53	Topic 54	Topic 55
circuit	mixture	organic	porous	assembly
integrated	precursor	lightemitting	inorganic	distal
memory	silicon	compound	water	tissue
external	preparing	emitting	binder	member
circuits	powder	electron	coating	tube
internal	slurry	efficiency	particles	proximal
chip	chemical	light	composite	jaw
function	cerium	derivative	separator	disposed
test	polishing	transporting	electrochemical	members
multiple	resulting	high	pores	handle

Topic 56	Topic 57	Topic 58	Topic 59	Topic 60
power	compound	lower	acid	compounds
voltage	producing	side	ester	treatment
control	reaction	plate	amino	relates
circuit	derivative	part	carboxylic	methods
output	reacting	upper	acids	cancer
supply	presence	parts	fatty	compositions
signal	intermediate	peripheral	alcohol	activity
switch	amine	fixing	derivative	comprising
input	produced	case	amide	diseases
level	production	collecting	salt	pharmaceutical

Topic 61	Topic 62	Topic 63	Topic 64	Topic 65
target	film	group	phase	sheet
nanoparticles	substrate	represents	gas	set
methods	thin	atom	fiber	glass
binding	spacer	alkyl	partially	greater
nanoparticle	tft	hydrogen	fibers	reduction
protein	counter	formula	deposited	equal
sample	thickness	represented	wide	band
biological	shielding	independently	deposition	increase
detection	films	substituted	separation	reduced
modified	substr	represent	thick	increased

Topic 66	Topic 67	Topic 68	Topic 69	Topic 70
light	weight	composition	polymer	drive
source	component	controlling	monomer	cleaning
emitting	parts	comprising	polymerizable	robot
wavelength	molecular	excellent	polymerization	system
reflective	mass	skin	comprising	motor
backlight	composition	active	polymers	control
color	rubber	plant	copolymer	object
led	amount	control	composition	chassis
sources	content	hair	conjugated	autonomous
reflected	based	shows	crosslinking	directed

Topic 71	Topic 72	Topic 73	Topic 74	Topic 75
cathode	top	devices	compositions	assembly
anode	bottom	materials	moiety	member
core	channel	electronic	therapeutic	position
comprising	forms	crystalline	polymeric	coupled
subject	elastomeric	methods	relates	disposed
electrochemical	surface	matrix	agents	mechanism
virus	recess	capable	methods	actuator
infection	forming	films	delivery	spring
discloses	extending	quantum	improved	locking
cycle	mold	applications	drug	guide

Topic 76	Topic 77	Topic 78	Topic 79	Topic 80
type	signal	optionally	salt	excellent
substance	system	ring	acceptable	composition
combination	applying	substituted	disease	resistance
addition	selectively	aromatic	pharmaceutic ally	properties
reducing	artifacts	heterocyclic	diseases	thermal
included	estimation	compound	treatment	heat
improving	noise	nitrogen	treating	curable
products	methods	substituents	effective	stability
contrast	algorithm	membered	preventing	adhesive
substances	related	phenyl	agent	good



3. ARM 결과

조건부		결론부	지지도	신뢰도	향상도
{topic01}	=>	{topic09}	0.0301	0.5430	5.2308
{topic02}	=>	{topic71}	0.0213	0.2513	5.7933
{topic02}	=>	{topic20}	0.0216	0.2555	4.4284
{topic02}	=>	{topic40}	0.0296	0.3493	4.4129
{topic02}	=>	{topic26}	0.0201	0.2377	2.7320
{topic04}	=>	{topic75}	0.0216	0.3664	4.5975
{topic04}	=>	{topic39}	0.0242	0.4114	4.7470
{topic09}	=>	{topic01}	0.0301	0.2902	5.2308
{topic09}	=>	{topic18}	0.0364	0.3506	5.0175
{topic09}	=>	{topic62}	0.0332	0.3200	4.6082
{topic09}	=>	{topic26}	0.0216	0.2085	2.3961
{topic12}	=>	{topic78}	0.0206	0.2390	3.3477
{topic12}	=>	{topic63}	0.0421	0.4882	3.4430
{topic12}	=>	{topic24}	0.0218	0.2533	1.9667
{topic17}	=>	{topic22}	0.0234	0.2752	5.7575
{topic17}	=>	{topic66}	0.0283	0.3323	4.9752
{topic18}	=>	{topic62}	0.0231	0.3312	4.7699
{topic18}	=>	{topic26}	0.0229	0.3274	3.7627
{topic18}	=>	{topic09}	0.0364	0.5209	5.0175
{topic20}	=>	{topic02}	0.0216	0.3752	4.4284
{topic22}	=>	{topic17}	0.0234	0.4898	5.7575
{topic24}	=>	{topic60}	0.0267	0.2071	2.6673
{topic24}	=>	{topic79}	0.0346	0.2689	3.0927
{topic24}	=>	{topic35}	0.0355	0.2757	2.5750
{topic24}	=>	{topic63}	0.0450	0.3491	2.4620
{topic26}	=>	{topic18}	0.0229	0.2629	3.7627
{topic26}	=>	{topic02}	0.0201	0.2315	2.7320
{topic26}	=>	{topic53}	0.0354	0.4071	4.6173
{topic26}	=>	{topic09}	0.0216	0.2487	2.3961
{topic30}	=>	{topic75}	0.0213	0.4064	5.0999

{topic30}	=>	{topic39}	0.0210	0.4013	4.6309
{topic34}	=>	{topic36}	0.0208	0.4417	7.5642
{topic35}	=>	{topic74}	0.0223	0.2079	3.7837
{topic35}	=>	{topic60}	0.0433	0.4043	5.2061
{topic35}	=>	{topic79}	0.0444	0.4150	4.7740
{topic35}	=>	{topic63}	0.0260	0.2426	1.7107
{topic35}	=>	{topic24}	0.0355	0.3317	2.5750
{topic36}	=>	{topic34}	0.0208	0.3555	7.5642
{topic38}	=>	{topic75}	0.0299	0.5007	6.2837
{topic38}	=>	{topic39}	0.0219	0.3674	4.2392
{topic39}	=>	{topic30}	0.0210	0.2426	4.6309
{topic39}	=>	{topic55}	0.0251	0.2895	5.4253
{topic39}	=>	{topic38}	0.0219	0.2528	4.2392
{topic39}	=>	{topic04}	0.0242	0.2793	4.7470
{topic39}	=>	{topic75}	0.0322	0.3710	4.6562
{topic40}	=>	{topic47}	0.0204	0.2578	3.9013
{topic40}	=>	{topic02}	0.0296	0.3739	4.4129
{topic47}	=>	{topic40}	0.0204	0.3088	3.9013
{topic50}	=>	{topic67}	0.0256	0.3806	5.1714
{topic50}	=>	{topic80}	0.0223	0.3307	4.5594
{topic53}	=>	{topic66}	0.0202	0.2295	3.4355
{topic53}	=>	{topic26}	0.0354	0.4018	4.6173
{topic53}	=>	{topic63}	0.0249	0.2826	1.9927
{topic53}	=>	{topic24}	0.0246	0.2786	2.1625
{topic55}	=>	{topic75}	0.0242	0.4536	5.6927
{topic55}	=>	{topic39}	0.0251	0.4702	5.4253
{topic57}	=>	{topic63}	0.0206	0.3467	2.4452
{topic57}	=>	{topic24}	0.0231	0.3899	3.0268
{topic60}	=>	{topic74}	0.0203	0.2617	4.7616
{topic60}	=>	{topic79}	0.0343	0.4414	5.0776
{topic60}	=>	{topic35}	0.0433	0.5575	5.2061
{topic60}	=>	{topic24}	0.0267	0.3436	2.6673
{topic62}	=>	{topic18}	0.0231	0.3333	4.7699

{topic62}	=>	{topic09}	0.0332	0.4784	4.6082
{topic63}	=>	{topic78}	0.0480	0.3383	4.7394
{topic63}	=>	{topic12}	0.0421	0.2966	3.4430
{topic63}	=>	{topic79}	0.0296	0.2087	2.4009
{topic63}	=>	{topic24}	0.0450	0.3171	2.4620
{topic66}	=>	{topic17}	0.0283	0.4233	4.9752
{topic66}	=>	{topic53}	0.0202	0.3029	3.4355
{topic67}	=>	{topic50}	0.0256	0.3481	5.1714
{topic67}	=>	{topic69}	0.0224	0.3037	4.1370
{topic69}	=>	{topic67}	0.0224	0.3045	4.1370
{topic69}	=>	{topic63}	0.0209	0.2852	2.0113
{topic71}	=>	{topic02}	0.0213	0.4908	5.7933
{topic74}	=>	{topic60}	0.0203	0.3698	4.7616
{topic74}	=>	{topic35}	0.0223	0.4051	3.7837
{topic75}	=>	{topic30}	0.0213	0.2672	5.0999
{topic75}	=>	{topic55}	0.0242	0.3038	5.6927
{topic75}	=>	{topic38}	0.0299	0.3747	6.2837
{topic75}	=>	{topic04}	0.0216	0.2705	4.5975
{topic75}	=>	{topic39}	0.0322	0.4035	4.6562
{topic78}	=>	{topic12}	0.0206	0.2884	3.3477
{topic78}	=>	{topic79}	0.0245	0.3428	3.9435
{topic78}	=>	{topic35}	0.0212	0.2970	2.7740
{topic78}	=>	{topic63}	0.0480	0.6720	4.7394
{topic78}	=>	{topic24}	0.0246	0.3441	2.6711
{topic79}	=>	{topic78}	0.0245	0.2815	3.9435
{topic79}	=>	{topic60}	0.0343	0.3943	5.0776
{topic79}	=>	{topic35}	0.0444	0.5112	4.7740
{topic79}	=>	{topic63}	0.0296	0.3404	2.4009
{topic79}	=>	{topic24}	0.0346	0.3984	3.0927
{topic80}	=>	{topic50}	0.0223	0.3069	4.5594
{topic35, topic60}	=>	{topic79}	0.0237	0.5469	6.2915
{topic35, topic79}	=>	{topic60}	0.0237	0.5328	6.8610
{topic60, topic79}	=>	{topic35}	0.0237	0.6907	6.4507

4. LDA R 코드

```
library("tm")
library("topicmodels")

#텍스트읽어오기
lgtext <- readLines("lgpatent.txt")

#코퍼스만들기
lgcorpus <- Corpus(VectorSource(lgtext))

#전처리
lgcorpus <- tm_map(lgcorpus, stripWhitespace)
lgcorpus <- tm_map(lgcorpus, tolower)
lgcorpus <- tm_map(lgcorpus, removePunctuation)
lgcorpus <- tm_map(lgcorpus, removeNumbers)
mystopwords <- c(stopwords("english"),stopwords("SMART"))
lgcorpus <- tm_map(lgcorpus, removeWords, mystopwords)
lgcorpus <- tm_map(lgcorpus, stemDocument)
lgcorpus <- tm_map(lgcorpus, PlainTextDocument)

#document - term matrix
lgdtm <- DocumentTermMatrix(lgcorpus)

#빈도수높은 불용어 제거
```

```

lgstopword <- as.matrix(lgdtm)
str(lgstopword)
lgstopword <- sort(colSums(lgstopword), decreasing=T)
lgstopword <- as.matrix(lgstopword)
head(lgstopword,30)
lgfilter <- (lgstopword>=898)
lgfilter <- cbind(lgstopword,lgfilter)
colnames(lgfilter) <- c("freq",'logic')
lgfilter <- data.frame(lgfilter)
lgfilter <- lgfilter[lgfilter[,2]==1,]
write.csv(lgfilter,"lgstopword.csv") #빈도수 높은 단어 색출

#stopword추가
mylgstopword <-
c(stopwords("english"),stopwords("SMART"),"first","second","one","can","i
nvention","may","two","method","also","provided","present","provide","infor
mation","least","comprise","back","selected","include","state","predetermined
","thereby","produce","associated","including","positioned","includes","produ
ct","process","formed","form","sufficient","end","use","method","preferably","
device","supports","provides","structure","provides","along","using","user","
material","made","functional","users","substantially","supported","used","con
figured","support","embodiment","applied","third","regions","region","thereof
","fourth","apparatus","easily","generally","effect","apart","allowing","current
","three","said","wherein")

lgcorpus <- tm_map(lgcorpus, removeWords, mylgstopword)

```

```
lgdtm <- DocumentTermMatrix(lgcorpus)
```

```
#lda
```

```
lglda <- LDA(lgdtm,80,method="Gibbs",control=list(iter=2000,seed=100))
```

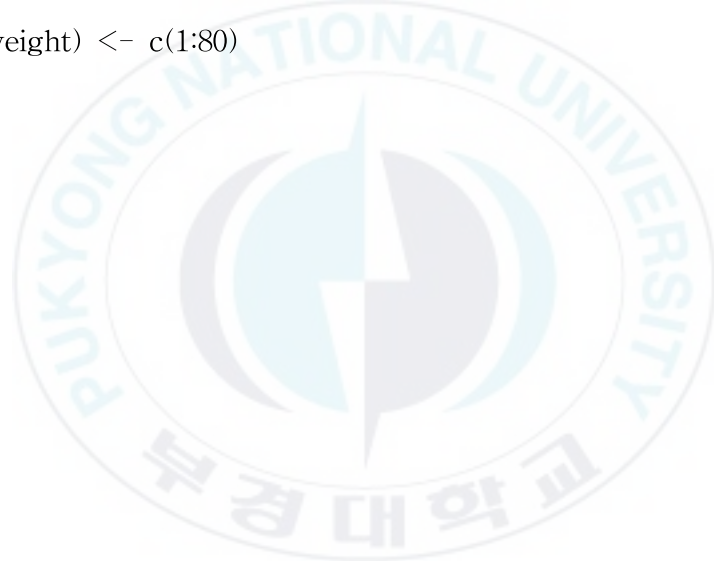
```
lgldaresult <- terms(lglda,10)
```

```
#문서별 토픽 가중치
```

```
lgweight <- as.data.frame(lglda@gamma)
```

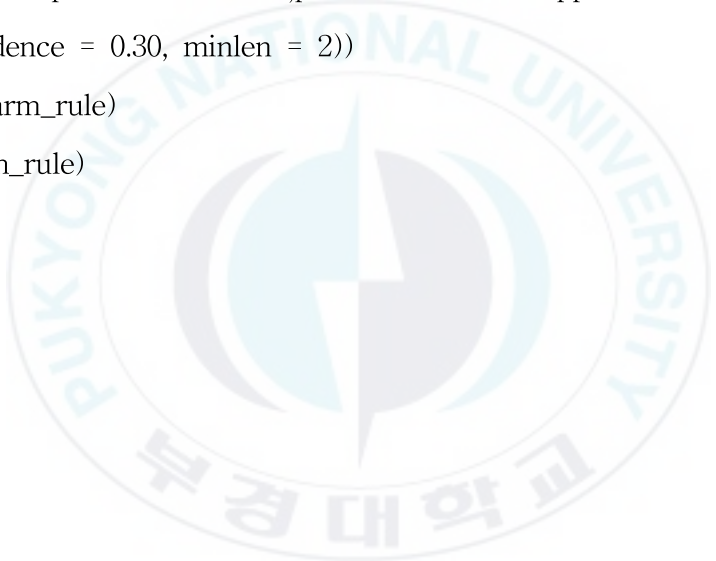
```
names(lgweight) <- c(1:80)
```

```
lgweight
```



5. ARM R 코드

```
library(arules)
arm<-read.csv("weightdata.csv",header = TRUE)
str(arm)
arm <- as.matrix(arm)
arm <- as(arm,"transactions")
arm_rule <- apriori(data = arm,parameter = list(support =
0.015,confidence = 0.30, minlen = 2))
summary(arm_rule)
inspect(arm_rule)
```

A large, light blue watermark logo of Pukyong National University is centered in the background. It features a circular emblem with a stylized 'P' and 'K' inside, surrounded by the university's name in English 'PUKYONG NATIONAL UNIVERSITY' and Korean '부경대학교'.

6. 유사도 계산 python 코드

```
>>> import logging
```

```
>>> logging.basicConfig(format='%(asctime)s : %(levelname)s :  
%(message)s', level=logging.INFO)
```

```
>>> from gensim import corpora, models, similarities
```

```
>>> documents = open('c:/lgsim.txt')
```

```
>>> texts = [[word for word in document.lower().split() ] for document  
in documents]
```

```
>>> dictionary = corpora.Dictionary(texts)
```

```
>>> dictionary.save('finalgsim.dict') # store the dictionary, for future  
reference
```

```
>>> corpus = [dictionary.doc2bow(text) for text in texts]
```

```
>>> corpora.MmCorpus.serialize('finalgsim.mm', corpus)
```

```
>>> from gensim import corpora, models, similarities
```

```
>>> dictionary = corpora.Dictionary.load('finalgsim.dict')

>>> corpus = corpora.MmCorpus('finalgsim.mm')

>>> lsi = models.LsiModel(corpus, id2word=dictionary, num_topics=80)

>>> doc = open('c:/ussim.txt')

>>> texts2 = [[word for word in document.lower().split() ] for
document in doc]

>>> cor = [dictionary.doc2bow(text) for text in texts2]

>>> vec_lsi = lsi[cor] # convert the query to LSI space

>>> index = similarities.MatrixSimilarity(lsi[corpus])

>>> index.save('finalgsim.index')

>>> index = similarities.MatrixSimilarity.load('finalgsim.index')

>>> import pprint

>>> pp = pprint.PrettyPrinter(indent=4)
```

```
>>> sims = index[vec_lsi]
```

```
>>> pp.pprint(list(enumerate(sims)))
```

