

# 텍스트 마이닝을 통한 핀테크 연관 핵심 기술 특허 추출 방법

이황로, 최은미\*

국민대학교 금융정보보안학과

\*국민대학교 비즈니스IT전문대학원

e-mail:rosa0615@kookmin.ac.kr emchoi@kookmin.ac.kr

## Essential Technical Patent Extraction Method Associated with Fintech Based on Text Mining

Hwangro Lee, Eunmi Choi\*

Graduate school of Financial Information Security, Kookmin University

\*Graduate school of Business IT, Kookmin University

### 요 약

금융과 IT가 융합되는 핀테크(Fintech)가 IT산업과 금융산업에 새로운 패러다임으로 급부상하고 있다. 핀테크 기술에 대한 기술동향을 파악하고 유사한 연관 기술을 도출하는 것은 관련 사업자가 시장 경쟁에서 우위를 차지하기 위해 필요한 전략적 방향을 제시해 준다. 하지만 핀테크와 같이 단 기간 내에 기술에 대한 과급 속도가 빠르게 일어나며 산업전반에서 기술선점의 필요성이 크게 대두되는 경우 특허 데이터베이스만으로 유사기술을 검색을 위한 키워드를 선정하는 것이 어렵다는 단점이 있다. 본 논문에서는 새롭게 이슈화되는 기술 중 그 성장세가 급격하게 변화하여 등록된 특허만으로는 연관 기술 영역을 파악하는 일이 번거로운 상황에서 기사 분석을 통해 연관 기술 키워드를 추출 할 수 있는 방법을 제안하고자 한다. 특히 핀테크에서 중요하게 인식되는 결제, 보안, 사용자환경에 대한 연관 기술 키워드를 기사 내용에 포함되는 단어의 빈도 분석을 통해 추출하고자 하였다. 최종적으로 추출된 기술 키워드를 이용하여 실제 특허 검색 데이터베이스에서 관련 특허를 수집하고 분석하여 핀테크와 관련성이 매우 높은 연관 핵심 기술 특허를 도출하였다.

### 1. 서론

핀테크에 대한 수요는 금융산업 전반에 걸쳐서 높은 상승세로 서비스 산업과 소비자의 지불 패턴의 변화가 일어날 것을 예견하고 있다. 2014년 하반기부터 ‘핀테크’를 중심으로 금융 산업 전반에서 패러다임의 변화가 급변하고 있다. 2014년 전반기 모바일 시장이 최근 몇 년 동안 엄청나게 큰 규모로 성장하였고, 기존 결제 방식의 복잡한 절차를 간소화 하려는 사용자의 요구가 커져갔다. 이러한 핀테크 기술들이 금융 산업의 전반에 융합되기 시작하면서, 빅데이터와 IT플랫폼을 통한 새로운 형태의 혁신적 금융 서비스가 등장하게 되고, 또한 기존의 금융 시스템들이 가져왔던 문제들을 개선하는데 크게 기여할 것으로 기대되고 있다.[1]

급변하는 기술에 대한 산업영역에서의 동향을 파악하기 위한 여러 방법 중 적절한 키워드를 통해 특허 데이터베이스를 검색하면 정량적인 수치와 해당 특허 사업자의 특허 포트폴리오 작성이 가능하다. 또한 수집된 자료로부터 데이터 분석 기법을 활용하여 동종 기술 사업자의 방향성이나 전략을 예측할 수 있으며, 공백기술영역을 파악하여

경쟁에서 우위를 차지할 수 있도록 한다. 하지만 최신 기술의 유사기술을 파악하기 위한 적절한 키워드 추출은 많은 시간과 인적자원을 필요로 하며, 이러한 자원은 사업자의 이익에 많은 손실을 가져오게 된다. 데이터 분석 기법 중 텍스트 마이닝은 구조화되지 않은 발문치에서 유용한 정보를 획득하는 방법으로 많은 주목을 받고 있으며 텍스트 마이닝 응용 프로그램은 다양한 영역에서 활용 될 수 있다.[2] 본 논문에서는 새롭게 이슈화되는 기술 중 그 성장세가 급격하게 변화하여 등록된 특허 정보만으로는 연관 핵심 기술 영역을 파악하는 일에 많은 자원을 소모해야하는 상황에서, 최신 기술 동향을 반영할 수 있는 기사 데이터를 수집하고 이를 텍스트 분석 기법을 통해 연관 기술 키워드를 추출하여 연관된 특허를 추출할 수 있는 방법을 제안하였다.

### 2. 관련 연구

본 장에서는 최신 기술에 대한 키워드를 추출하기 위해 데이터 수집에 필요한 웹 크롤러 그리고 수집된 정보로부터 연관 기술 특허를 도출하기 위한 분석 기법을 소개한다.

#### 2.1 웹 크롤러(Web Crawler)

웹 크롤러(web crawler)는 조직적, 자동화된 방법으로 웹

\* 교신저자

이 논문은 2013년 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임. (No. 2011-0011507)

드 와이드 웹을 탐색하는 컴퓨터 프로그램이다.[7] 다양한 프로그래밍 언어에서 웹 크롤링을 위한 라이브러리를 제공하고 있으며 대표적으로 사용되는 웹 크롤러를 지원하는 라이브러리는 다음과 같다.

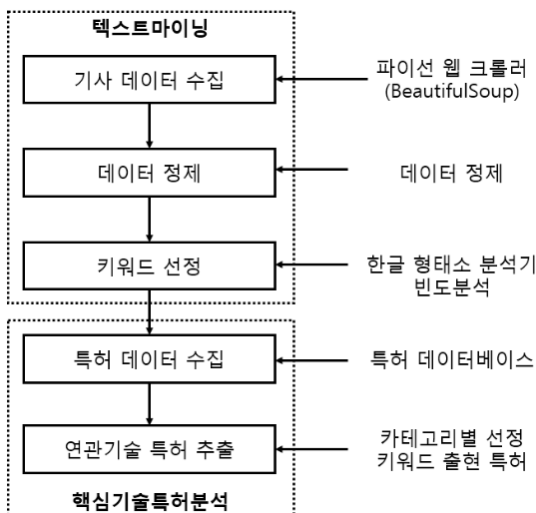
- **BeautifulSoup** : Python에서 사용하는 HTML 파서이다. HTML과 XML 파일로부터 데이터를 추출하기 위한 Python 라이브러리이다.[8]
- **jsoup** : JAVA에서 사용하는 HTML 파서이다. jsoup는 HTML 문서를 읽어들이어 문서를 DOM 객체로 변환을 하고 jsoup에서 지원하는 selector api를 이용하여 특정 Element에 접근할 수 있게 하며, 해당 Element의 정보를 읽거나 수정할 수 있다.

### 2.2 텍스트 마이닝

텍스트 마이닝(Text Mining)은 대용량의 데이터에서 사용자가 관심을 가지는 정보를 키워드의 수준 뿐만 아니라 Context의 수준의 의미를 찾아내는 프로세스를 의미한다.[3] 웹 상의 정보가 폭발적으로 증가하고 특히 비정형 데이터가 증가함에 따라 이러한 데이터를 처리할 수 있는 방법이 필요하게 되었으며, 데이터 속에서 패턴을 발견하고 특정 주제와 연관된 데이터를 검색하는 방법으로 점점 발전하고 있다. 텍스트 마이닝의 일반적인 처리 절차는 1) 비정형 정보수집 2)정보처리 3)정보추출 4)정보분석의 4가지 절차를 따르고 있다. 수학적 모델이나 알고리즘을 적용하여 대용량의 비정형 데이터에서 유용한 정보를 추출하거나 유사한 문서를 집단으로 구분하기 위한 방법으로 많이 사용되어 진다.

### 3. 핵심 기술 특허 추출 방법론

본 연구에서 제시하는 핀테크 연관 기술 키워드 추출 방법은 아래 (그림 1)에서 제시된 바와 같이 크게 두 개의 세부 방법론을 진행하며, 기사를 통한 텍스트마이닝 부분과 핵심기술특허분석 부분으로 나누어진다.



(그림 1) 연관기술 키워드 분석 과정

### 3.1 기사 데이터 수집

첫 번째 단계는 데이터를 수집하는 단계로서, 먼저 핀테크 연관 기술 키워드 추출을 위해 기사 데이터를 수집한다. N언론사의 2014년 8월부터 2015년 7월까지 1년간의 ‘핀테크’단어 출현 기사를 수집하였으며 총 980건의 기사에 대하여 작성일, 제목, 기자, 기사내용에 대한 데이터를 수집하였다.

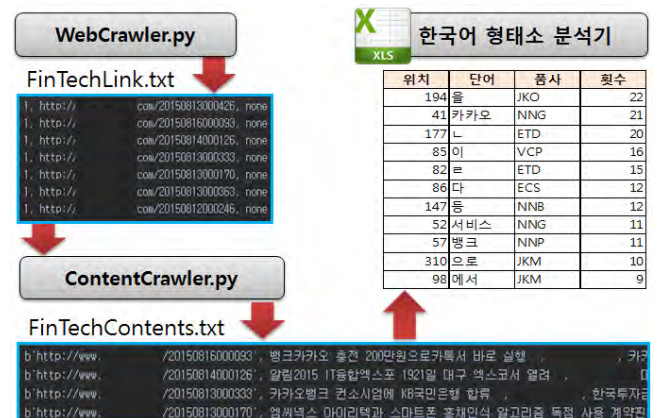
<표 1> N언론사 월별 핀테크 연관 기사 수

년	2014년					2015년						
	8	9	10	11	12	1	2	3	4	5	6	7
기사 수	2	7	25	18	60	95	111	165	127	122	107	98

각 월별 핀테크 연관 기사 수는 <표 1> 과 같다. 2014년 10월부터 핀테크에 대한 관심이 큰 폭으로 증가하기 시작하였고 2015년 3월에는 165건의 기사가 게재될 만큼 단기간 내에 빠르게 이슈화 되고 있음을 확인할 수 있었으며 2015년 7월까지 지속적으로 언급되고 있음을 알 수 있다.

### 3.2 텍스트마이닝을 통한 데이터 정제

두 번째 단계는 자연어로 구성된 980건의 기사 본문에서 키워드를 추출하고, 이를 정규화 하고 정제하는 단계이다. 언론 기사 데이터 수집을 위해 Python BeautifulSoup4 라이브러리를 이용하였으며 기사 url을 먼저 수집하는 단계 거쳐 데이터 수집이 이루어졌다.



(그림 2) 데이터 정제 정규화 과정

데이터 수집으로부터 데이터 정제 단계는 (그림 2)과 같으며 세부적인 절차는 다음과 같다. 1) WebCrawler.py 스크립트를 이용하여 기사에 대한 URL을 FinTechLink.txt에 저장한다. 2) ContentCrawler.py 스크립트를 이용하여 FinTechLink.txt의 URL을 불러온다. 3) BeautifulSoup4 패키지를 이용하여 날짜, 제목, 기자, 내용 데이터를 추출하여 FinTechContents.txt에 구분자 ‘,’와 같이 CSV 형태로 저장한다.

### 3.3 키워드 선정

세 번째 단계는 정제된 데이터로부터 한국어 형태소 분석기를 이용하여 기사에 대한 빈발 단어를 추출하는 단계이다. 특허데이터베이스 검색식에 사용될 핀테크 관련 키

워드 추출은 자연어 처리를 위한 다양한 모듈 및 자료 구축을 가능하게 해주는 꼬꼬마 형태소 분석기를 사용하였다.[5] 빈도 분석을 통해 수행되었으며 총 980건의 기사에서 각각 결제, 보안, 사용자 환경에 대한 기사를 추출하기 위해 기사 제목에 결제, 보안, 사용자 환경을 포함하는 기사를 검색하여 총 3개의 데이터 셋을 구성하였다. 결제와 관련된 데이터 셋은 총 187건이며, 보안과 관련된 데이터 셋은 112건, 사용자 환경과 관련된 데이터 셋은 25건이 추출되었다.

<표 2> 핀테크 관련 기사 단어 빈도 분석

결제관련 기사 단어빈도			보안관련 기사 단어빈도			사용자환경 관련 기사 단어빈도		
단어	형태	빈도	단어	형태	빈도	단어	형태	빈도
ㄴ	ETD	2672	ㄴ	ETD	1657	ㄴ	ETD	294
를	JKO	2375	을	JKO	1641	를	JKO	234
다	ECS	2126	다	ECS	1300	다	ECS	213
를	JKO	1881	를	JKO	1131	를	JKO	182
에	JKM	1664	에	JKM	993	이	JKS	179
10,720개의단어출현			7,913개의단어출현			2,648개의단어출현		
독립적	NNG	1	열	VV	1	연사	NNG	1
연신금융법	NNG	1	innovation	OL	1	초청	NNG	1

형태소 분석기로 분석한 결과 다양한 형태의 단어들이 출현한 것을 <표 2>을 통해 확인할 수 있다. 또한 <표 3>과 같이 핀테크 관련 키워드를 확인하기 위해서 우선, 단어의 출현빈도와 관계없이 명사와 외국어를 제외한 품사들을 제하였다.

<표 3> 핀테크 관련 기사 명사 추출

결제기사 명사추출		보안기사 명사추출		사용자환경 기사 명사추출	
단어	빈도	단어	빈도	단어	빈도
핀	1434	핀	925	핀	152
핀테크	1328	금융	916	금융	148
금융	1278	핀테크	843	핀테크	144
결제	744	보안	365	은행	75
서비스	618	결제	359	서비스	71
7,512개의명사추출		5,629개의단어출현		1,853개의단어출현	

핵심 유사 키워드 빈발빈도가 높은 단어 중에 각각 결제, 보안, 사용자환경과 밀접한 연관이 있는 단어들로 선정하였다. 이러한 절차를 거쳐 최종적으로 <표 4>에서 보는 것과 같이 핀테크 관련 핵심유사키워드를 추출할 수 있다

<표 4>핀테크 관련 핵심 유사 키워드 추출

결제관련 빈발단어 Top10		보안관련 빈발단어 Top10		사용자환경 빈발단어 Top10	
단어	빈도	단어	빈도	단어	빈도
금융	1278	보안	365	서비스	71
결제	744	카드	267	고객	38
은행	472	인증	136	통신	27
전자	399	생체	97	경험	19
투자	233	토큰	71	플랫폼	17
거래	103	해킹	31	속도	15
기관	96	탐지	26	인프라	11
외환	57	악성	25	장비	10
주식	49	방지	20	간편	8
대금	45	생성기	16	음성	7

#### 4. 핵심 기술 특허 추출

##### 4.1 특허 데이터 수집

네 번째 단계는 도출된 키워드로부터 특허 데이터베이스를 이용하여 검색식을 작성하고 핀테크와 관련된 모든 특허데이터를 수집하는 단계이다.

특허 DB 검색식을 구성하기 위해 본 연구에서는 다양한 유·무료 특허 데이터베이스 중 KIPRIS (<http://www.kipris.or.kr>)와 WIPSON(<http://wipson.com>)을 이용하였다. KIPRIS, WIPSON은 온라인 검색을 통해 특허문서를 검색하여 편리하고 빠르게 특허를 수집할 수 있도록 지원해주는 서비스이다. 또한 여러 부가 기능을 통해 사용자가 원하는 결과를 검색할 수 있도록 고급검색과 다양한 부가서비스를 제공한다. 특허 데이터베이스를 이용하여 기사분석을 통해 도출된 핵심 유사 키워드를 검색어에 이용하여 특허를 검색하고 분석을 위한 핵심 정보를 수집하였다.

##### 4.2 연관 기술 특허 추출

다섯 번째 단계는 수집된 특허 데이터를 전처리하는 과정이 이루어졌고 연관 기술 특허 추출을 위한 테이블을 작성하였다. 특허DB 생성을 위한 핵심특허 도출 작업은 관련성이 적은 특허를 제거하고 핵심특허를 추출하기 위해 수행되어 졌으며 아래 사항에 해당하는 특허를 제외한 모든 특허는 제거 하여 분석을 수행하였다.

- 1) 최근 10년간의 특허데이터를 추출 (2005년~2015년)
- 2) 발명의 명칭에 유사 키워드가 포함된 특허만을 추출
- 3) 출원되어 발명의 명칭이 공개된 특허데이터를 추출
- 4) 중복된 특허 제거

핵심특허 도출을 통한 특허 추출을 수행하여 <표 5>과 같이 최종 선정특허를 수집하였다. 발명의 명칭에 대한 중요 가중치가 가장 높을 것으로 판단하여 수집한 특허 데이터에서 요약정보와 제1청구항의 정보를 제외시켰다.

<표 5> 핵심 유사 키워드를 이용한 특허 수집

출원일	발명의 명칭	결제 유사키워드 유무	보안 유사키워드 유무	사용자 환경 유사키워드 유무
2015-07-28	휴대폰을 이용한 결제	1	0	1
2015-07-20	휴대폰의 프로그램을	1	0	1
2015-07-14	결제수단 등록을 위한	1	1	1
2015-05-26	이종의 디스플레이 간	1	0	1
2015-04-22	공간위치기반 카드결제	1	0	0
2015-04-16	단말기 할부 청구금액	0	0	0
2013-12-27	무선 통신 단말 및 이	0	0	1
2013-12-27	무선 통신 단말 및 이	0	0	0
2013-12-03	카드 안심 결제방법(A	1	1	0
2013-10-25	서비스 제공 장치와 P	1	0	0
2013-09-23	서버, 단말 장치 및 그	0	0	1
2013-08-21	경로 안내 서비스 중	0	0	0
2013-08-07	이동전화 애플리케이션	1	0	0

기술의 성장은 기술이 속한 환경 속에서 정치, 경제, 사

회적 영향을 받기 때문에 효과적인 기술예측을 위해서 먼저 예측대상기술에 대한 충분한 이해가 필요하다.[4] 또한 대상기술과 경쟁관계에 있는 유사기술에 대한 분석도 수반되어야 한다. 그러므로 수집된 특허DB로부터의 핵심특허 도출은 이러한 상황을 종합적으로 고려하여 대상기술과 기술이 적용될 환경에 맞게 수행되어야 한다.

### 5. 핵심 기술 특허 도출 결과

최종 선정된 핀테크 기술관련 특허 건수는 <표 6>에서 보는바와 같이 KPRIS 데이터베이스에서 결제 61건, 보안 9건, 사용자 환경 57건이고 WIPS 데이터베이스에서 결제 70건, 보안 9건, 사용자환경 63건 이다.

<표 6> 핀테크 기술 관련 특허 건수

분류	특허 데이터 다운로드 수 (KPRIS)	핵심특허 도출건수 (KPRIS)	특허 데이터 다운로드 수 (WIPS)	핵심특허 도출건수 (WIPS)
결제	198 건	81 건	341 건	70 건
보안	38 건	9건	61 건	9 건
사용자 환경	210 건	57건	274 건	63 건

<표 7> 핀테크 기술 관련 특허 건수(2분류 중복특허)

분류	핵심특허 도출건수 (KPRIS)	핵심특허 도출건수 (WIPS)
결제 + 보안	9 건	11 건
보안 + 사용자환경	8 건	3 건
결제 + 사용자환경	33건	24 건

<표 8> 핀테크 기술 관련 특허 건수(3분류 중복특허)

분류	핵심특허 도출건수 (KPRIS)	핵심특허 도출건수 (WIPS)
결제 + 보안 + 사용자 환경	4 건	2 건

도출된 핀테크 기술 관련 특허 건수에 대해 기술 분류별 (발명의 명칭에 각각 결제, 보안, 사용자환경에 대한 유사 키워드가 출현한 문서), 2분류 중복특허(발명의 명칭에 결제, 보안, 사용자환경 중 두 가지 유사키워드가 동시에 출현한 문서 <표 7>), 3분류 중복특허(발명의 명칭에 결제, 보안, 사용자환경에 대한 유사키워드가 모두 동시에 출현한 문서 <표 8>) 선정과정을 거쳐 도출되었다.

<표 9> 핀테크 연관 핵심 기술 특허(KPRIS)

발명의명칭
카드 안심 결제방법 (ASSURED CARD PAYMENT METHOD)
휴대폰을 이용한 거래 인증방법 (A METHOD FOR AUTHORIZATION OF A TRANSACTION WITH THE USE OF A MOBILE PHONE)
모바일 상거래 인증 및 인가 시스템 (A MOBILE COMMERCE AUTHENTICATION AND AUTHORIZATION SYSTEM)
모바일 카드의 금융 프로그램 관리 방법 및 시스템 (System and method for managing finance program in mobile card)

<표 10> 핀테크 연관 핵심 기술 특허(WIPSON)

발명의 명칭
금융 결제와 통신 기능이 통합된 모바일 카드, 이를 이용한 해외 금융 결제 시스템 및 그 방법(Mobile card integrated finance settlement and telecommunication function, and overseas finance settlement system using the mobile card, and method thereof)
이동통신 단말기의 외장형 변환 젠더와 이를 이용한 선불형 IC 카드 충전 시스템 및 금융 거래 시스템(EXTERIOR CONNECTOR GENDER OF MOBILE COMMUNICATION APPARATUS, AND SYSTEM FOR SAVING MONEY IN IC CARD AND MONEY TRANSACTION SYSTEM USING THE CONNECTOR GENDER)

이러한 절차를 거치게 되면 <표 9>, <표 10>과 같이 텍스트 마이닝 분석 관점에서 핀테크 연관 기술 특허와 관련성이 가장 높은 특허만을 추출할 수 있다.

### 6. 결론 및 향후 연구

본 논문에서는 새롭게 이슈화되는 기술에 대해 등록되어 있는 특허를 파악할 수 있도록 기사 분석을 통해 연관 기술 키워드를 추출 할 수 있는 방법을 제안하였다. 핀테크에서 중요하게 인식되는 결제, 보안, 사용자환경에 대한 연관 기술 키워드를 기사 내용에 포함되는 단어의 빈도분석을 통해 추출하였으며 추출된 기술 키워드를 이용하여 실제 특허 검색 데이터베이스에서 검색식을 이용하여 핀테크 특허 연관기술을 선정하였다. 이러한 연관기술 키워드 추출 방법은 앞으로 나올 기술의 핵심특허를 도출하는데 효과적으로 활용될 수 있다. 차후 연구주제로서 본 연구에서 제시한 방법을 통해 최종 선정된 핀테크 기술 관련 특허를 기반으로 PTDM(Patent-Term Matrix)생성과 대상 업체별 특허포트폴리오 작성, OS매트릭스 분석 등의 방법을 활용하여 심층적인 분석을 수행할 예정이다.

### 참고문헌

- [1] 김수형, 조영섭, 최대선, “핀테크 시대: 새로운 인증 기술을 요구하다”, 정보과학회지, 33권, 5호, pp. 17-22, 2015.
- [2] Heeyong Noh, Yeongran Jo, Sungjoo Lee, “Keyword Selection and Processing Strategy for Applying Text Mining to Patent Analysis”, Export Systems with Applications, Elsevier, Vol 42, Issues 9, June 2015, pp4348-4360
- [3] 정근하, “텍스트마이닝과 네트워크 분석을 활용한 미래예측 방법 연구”, 한국과학기술기획평가원, Dec 2010
- [4] 전성해, 박상성, 장동식, “특허분석과 기술예측”, 교우사, 2014
- [5] 안정진, 최은미, “특허분석 기반의 핀테크 보안 기술 로드맵 수립”, 정보처리학회추계학술발표대회, 2015. (제출)
- [6] kkma.snu.ac.kr
- [7] ko.wikipedia.org/wiki/웹\_크롤러
- [8] www.crummy.com/software/BeautifulSoup/