

PATENT MAP 작성 자동화 방법

Automatic Patent Map Drawing

원정욱(J.W. Won)

이형복(H.B. Lee)

고재상(J.S. Koh)

IT기술분석팀 선임연구원

지적재산팀 책임행정원, 팀장

광통신부품연구센터 책임연구원, 센터장

특허지도는 엔지니어가 특정기술을 개발할 때 많은 도움을 준다. 그리고 프로젝트 관리자는 기술적 흐름과 시장상황을 한눈에 파악할 수 있어 도움을 받을 수 있다. 본 고에서는 특허맵의 동향과 자동적으로 특허지도를 작성하는 방법에 대하여 소개하고자 한다.

I. 서론

PM Tool은 특허를 실무로 하는 사람이면 거의 다 아는 용어이다. 특허가 연구소에서도 관심을 갖게 된 지 십 여년 정도 지났다. 이제는 특허관리를 질적으로 해야 하고 또한 사용자중심으로 변해야 될 시기가 되었다. 이러한 요구를 만족시켜줄 수 있는 시스템이 PM Tool이며 여기서 나온 결과물이 PM이다. 본 고에서는 현재 PM의 동향에 대하여 잠깐 살펴보고 PM Tool을 자동화 할 수 있는 시스템을 소개해 보고자 한다.

II. PM의 태동

Patent Map(이하 PM이라 한다)은 1968년 일본 특허청에 의해 발간된 “내일을 밝히는 특허”에서 기술동향의 표시가능성이 시사되었고 1971년 소련의 안뵈르스키에 의하여 출간된 “기술예측의 체문제”에서 특허정보의 초록카드작성 및 통계처리 가능성을 제시하였다. 1972년에 일본 선진 기업을 중심으로 PM 작성을 본격 검토하였고 러시아에 특허정보이용

조사를 위해 조사단을 파견하였다. IBM 특허 계약 부장 초청 강연회에서 화제로 부각되면서 1974년에 WIPO 주관으로 PM에 대한 심포지엄을 개최했고 1980년대 중반부터 국내에서는 여러 기업이 PM 기법을 도입해서 활용하기 시작하여 PM 경진대회를 개최하였다.

III. PM 정의

“Patent Map(특허 지도)”이란 특허정보의 서지적 사항(출원인, 각종 특허번호 등)을 분류 정리하고, 기술내용(특허청구범위, 발명의 상세한 설명, 도면의 간단한 설명 등)을 분석 가공하여 그 데이터를 가지고 여러 정보를 한눈에 알기 쉽게 읽을 수 있도록 도표화한 것이다. PM은 시각성을 높여 과거로부터 현재에 이르기까지의 기술 동향, 출원인 동향 및 기술분포 동향과 복잡하게 얽혀 있는 권리관계를 이해하기 쉽게 일목 요연하게 작성한 도표로서 특허정보를 여러 가지 방법으로 정리, 분석하면 과거의 문제점, 앞 사람이 개발하지 않은 공백영역, 현재기술로부터 발전된 가까운 미래의 예측 등을 살필 수 있는

것으로 “특허정보의 분석도”라고도 하며, 또한 “특허의 흐름 조사도”라고도 불리운다.

IV. PM 필요성

현대 산업사회에서 특허의 중요성은 두말할 필요도 없다. 우수한 특허를 만들기 위해서는 기술분야와 관련된 특허들의 경영정보적 측면, 기술정보적 측면, 권리보호적 측면에서 체계적으로 접근할 필요가 있다. 초기에는 새로운 아이디어라고 하면 무조건 출원하였으나 이를 한차원 높여 체계적으로 대응하는 방법을 생각해야 하며 그러기 위해서는 PM으로 만들어진 정보는 상기의 3가지 측면을 제공하기 위해 충분하다.

PM에서 가장 중요한 기능을 선정하라고 하면 기술정보적 측면의 선행기술조사이다. 일반 지도는 만들어진 길을 찾기 위해 사용되나 특허지도는 남이 가지 않은 길을 찾기 위해 만들어진 필요한 지도이다 라는 말은 매우 의미 심장한 말이라 할 수 있겠다. PM을 작성하여 기술개발동향을 파악하고 자기가 개발하고자 하는 기술을 파악해 중복투자 방지는 물론이요 관련있는 기존의 특허에서 많은 아이디어를 얻을 수 있다. 또한 기술적으로 미개발되었거나 틈새기술 등을 발견할 수 있다. 이를 표로 정리하여 보여 주기 때문에 사용자는 쉽게 자기기술의 위치 즉 쓸모없는 것인지, 중요한 것인지 원천특허인지 등을 쉽게 알 수 있어 PM은 기술정보적 가치가 크다고 할 수 있다.

PM을 경영정보적 측면에서 살펴보면 PM은 경영자 또는 관리자에게 경쟁업체의 동향을 제시할 수 있다. 경쟁업체가 어느 분야에 집중투자하고 있는지 어느 기술에 강한지를 알 수 있어 전략적으로 이에 대처할 수 있는 경영정보를 경영자는 제공받게 된다. 또한 어느 발명자가 유능한지를 알아 인재 스카우트의 자료로도 활용할 수 있다. PM을 권리보호 측면에서 접근하면 특허를 분석하여 구성요건과 특허취득 가능성을 파악할 수 있다.

V. PM 작성방법

1. 작성절차

PM Tool을 이용한 PM 작성프로세스를 크게 나누면 5단계로 나누어 볼 수 있다. 1단계는 목표설정 단계로서 정보를 수집하는 단계, 2단계는 데이터를 추출하는 단계, 3단계는 데이터를 분류 정리하는 단계, 4단계는 데이터를 분석하는 단계, 그리고 5단계는 데이터를 표시해주는 단계로 나누어 볼 수 있다.

1단계에서는 먼저 어느 용도로 사용할 것인지 결정한다. 기술흐름을 보려는 것인지 경쟁업체의 개발 동향을 파악하려는 것인지 아니면 선행기술을 조사하기 위한 것인지 목적을 분명히 한다.

2단계에서는 외부데이터베이스(KIPROIS, USPTO 등)를 접속하여 데이터를 다운받도록 한다. 이때 주의할 점은 만일 PM Tool이 인터페이스 기능이 좋아 다운받기가 용이하면 문제 없지만 접속에서 명령어가 각 데이터베이스 마다 조금씩 다르기 때문에 경우에 따라서는 엑셀로 다운받든지 아니면 당해 데이터베이스에 맞게끔 Query를 날려서 해당데이터를 받아야 한다.

3단계는 다운로드 받은 데이터 또는 엑셀로 다운로드 받은 데이터를 병합하기도 하고 중복된 데이터는 삭제하기도 하여야 한다. 모든 것이 자동화되면 좋지만 일부 수작업이 필요한 경우가 생긴다. 예를 들어 동일인인데도 표기가 다르든지(삼성전자와 삼성전자(주)) 하는 경우이다. 요즘 한국은 전자출원시스템이 잘 되어 있기 때문에 출원인 코드가 있어 구별이 용이하지만 이 시스템 이전이든지 외국특허의 경우에는 아직도 작업을 해주어야 하는 경우가 발생한다.

4단계는 데이터를 분석하는 것으로 필요한 것만 볼 수 있도록 옵션을 설정하는 경우가 대부분이다. 물론 전체를 다 보고자 하는 경우에는 특별한 옵션이 필요하지 않지만 다출원 국가 랭킹 3위까지 본다든지 하여 제약을 준다.

5단계는 데이터를 표시해주는 단계로 자동화로 데이터가 표현되어 지고 사용자의 요구에 따라 데이터 표현 형태도 변할 수 있게 할 수 있고 엑셀로 다

운로드 받아 더 세련된 표를 만들 수도 있다.

2. 검색사이트

특허검색을 웹을 이용하여 할 수 있는 국내외 사이트는 다음과 같다. 상기에서 설명한 바와 같이 아래 사이트에 연결하여 데이터를 다운로드 받을 수 있고 이를 통일화하기 위해서는 연동시스템이 필요하다.

가. 한국특허청 기술정보서비스(KIPRIS)
[<http://www.kipo.go.kr>]

특허검색을 클릭하거나, 특허기술정보센터 홈페이지[<http://www.kipris.or.kr>]로 찾아 가면 된다. KIPRIS는 국내 산업재산권 정보를 수집하여 종합적으로 구축한 온라인 산업재산권 정보서비스로 국내 특허, 상표검색이 가능하며 미국, 일본, 유럽특허의 검색을 위한 링크가 되어 있다. 20:00~08:00 사이는 무료이다.

나. WIPS의 세계특허검색서비스
[<http://www2.wips.co.kr>]

기존에 CD-ROM의 형태로 판매되던 한국지적재산권(특허, 실용신안, 상표, 의장)의 데이터를 CD-ROM이 아닌 인터넷상으로 서비스하여 R&D 및 관련부서에서 필요로 하는 정보를 쉽고 정확하게 찾아 볼 수 있도록 구성된 특허검색 인터넷 홈페이지이다. 한국 공개특허/실용신안, 상표, 의장은 물론 미국 특허/의장, 유럽 공개특허, 일본 공개특허/실용 등의 검색이 가능하며, 원문도 출력 가능하다. 일본 특허의 경우 한국어 검색이 가능하며, 동시 번역 서비스도 제공되고 있다. 회원가입이 필요하며 유료사이트이다.

다. 천리안 더웬트 특허검색
[<http://wpi.chollian.net>]

더웬트의 WPI(World Patent Index)에 대한 특허 정보를 그동안 상용 온라인 데이터베이스를 통하여 검색을 실시하였으나, 데이콤에서 시스템을 개발하여 웹을 통하여 세계 특허 정보를 검색할 수 있도록 하였다. 회원가입이 필요하며 유료사이트이다.

라. IBM 특허검색사이트[<http://www.delphion.com>]

IBM이 제공하는 특허 검색사이트이다. 1971년부터 1973년 사이의 일부 특허를 포함하여 1974년부터 현재까지 특허정보를 전산화하여 제공한다. 미국의 IBM사가 자사의 1971년 이후 등록특허를 모아놓은 사이트로서 검색 후 그 해당 원문까지 볼 수 있다(일주일마다 업데이트). 또한 일본특허의 영문 초록, 유럽특허, PCT 출원에 의한 국제특허 등도 검색이 가능하다. 다량의 원문이 필요한 경우 PDF 파일 형태로 주문하거나, 카피한 것을 우편으로 받을 수도 있다. 유료사이트이다.

마. 미국특허상표청 특허검색
[<http://www.uspto.gov>]

원문(full text)을 액세스하고 검색할 수 있는 기능을 무료로 이용할 수 있다. 1976년 이후 2백만 건 이상의 텍스트 데이터를 참고할 수 있다. 문서들의 full-page 문서 이미지까지 함께 볼 수 있다. 1976년부터 현재까지의(매주 갱신) 내용을 포함하고 있다.

바. QPAT-US[<http://www.qpat.com>]

1974년부터 미국특허를 서비스하고 있다. 유료 사이트로 2,000,000여 건을 보유하고 있고 \$195/월 \$1995/년으로 전문제공한다.

사. 유럽특허 사무국[<http://www.epo.co.at>]

유럽특허, 일본특허 영문초록, PCT 출원특허, 그리고 WIPO에 영문특허 요약을 제출하는 전세계 특허를 검색 가능하다. 그리고 각 국별 특허 사무국, 특허 관련 사이트 링크를 제공한다.

아. 일본특허 전자도서관
[<http://www.ipdl.jpomiti.go.jp/homepg.ipdl>]

PAJ로서 일본특허 공개정보를 제공하던 일본특허청 홈페이지로서 1993년 이후 특허의 공보 원본도 제공이 되고, 1993년 이후 프론트 페이지까지 검색이 가능하다. 일본특허가 많이 필요한 경우 대단히 유용한 사이트이다. 일본특허에 대한 타 정보 이

외에 단순히 일본특허의 영문초록 검색만을 위해서는 독일 PATON사가 제공하는 PAJ 검색사이트가 더 빠르고 편하다[URL: http://athena.patent-inf.tu-ilmeneau.de].

3. PM 종류

PM Tool을 이용해 나타낼 수 있는 맵의 종류는 <표 1>과 같다. PIAS 같은 특허작성시스템 틀은 성질에 반하지 않는 한 그림의 형태를 막대, 꺾은선, 곡이 등의 그래프로, 구성비율 등을 자유자재로 그래프를 변경하면서 표현할 수 있도록 되어 있다. 일반적으로 가장 많이 쓰이는 것은 구성비율을 보려는 것으로 출원인 비율, 국가간 비율 등을 나타내어 주는 맵들이다. 매트릭스 맵은 복수의 관점인 기술-응용분야-출원인 등의 삼면 관계를 나타내주는 데 주로 사용한다. 레이더 맵은 복수의 요소를 방사선으로 표시하여 기술분야-연도를 방사선으로 표시해 기술의 밸런스 기술분야별 주력도, 중점화 분야를 알 수 있다. 구성요소는 제품을 중심으로 각 구성마다 어떤 특허들이 있는지 표시하여 주는 특허맵이다.

<표 1> 입력방식에 따른 소요 대역폭

목적	기능	종류
- 기술개발 동향 파악 - 연구개발 테마 선정 - 틈새기술의 발견 - 기술과급상황	Patent Technical Map (기술정보적 활용 Map)	- 요지 Map - 레이더 Map - 기술분포 Map - 기술 Flow Map - 템페스트 Map - 매트릭스 Map - 기술상관 Map
- 경쟁기업 동향 파악 - 상품개발 흐름 파악 - 신 용도 개발 - 시장참여 상황 파악 - 기술적 범위 확인 - 발명자(인재) 동향 파악	Patent Management Map (경영정보적 활용 Map)	- 출원건수 동향 Map - 출원인 분포 Map - 발명자 Map - Patent Share Map - Ranking Map - New Entry Map - 기업상관 Map
- 기술적 범위 확인 - 특허취득 가능성 파악 - 특허망 형성	Patent Claim Map (권리정보적 활용 Map)	- Patent Family Map - 구성요건 Map - Claim Point Map - 권리관계 Map - 심사관계 Map

<자료>: 특허맵 개념 및 활용, 특허정, 2001. 3.

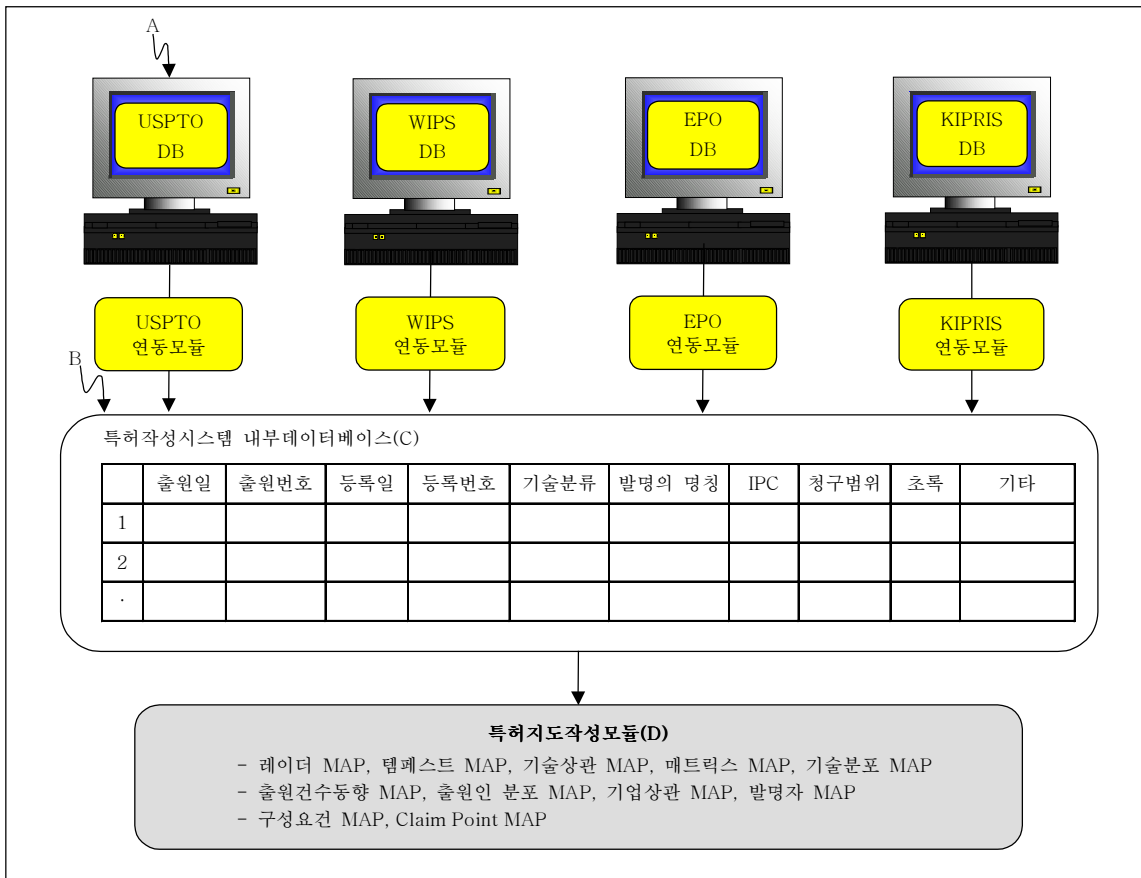
VI. PM 작성 자동화기법

기술분류에 의한 특허지도를 작성할 때 가장 큰 이슈는 특허지도를 손쉽게 작성하는 것이다. 기존에는 일일이 명세서 또는 초록을 보고 직접 기술분류를 하였으므로 시간이 많이 소요되었다. 본 고에서는 이러한 문제점을 극복하고 기술관련 단어를 손쉽게 추출하여 특허와 기술분류단어와 매칭시킴으로 기술분야관련 특허지도를 손쉽게 작성할 수 있도록 하였다. 간단히 설명하면 발명의 명칭, 청구범위, 초록, 발명의 상세설명 중 기술분야 단어를 추출할 대상을 선정 한 후 여기에서 불필요한 단어를 제거하고 필요한 데이터만을 뽑아내는 것이다. 이렇게 하기 위해서 불용 단어를 제거하는 기능이 필요하고 특허의 개수를 세어 각 단어의 가중치를 계산하는 방법을 사용하였다.

본 아이디어는 특허지도를 작성할 때 기술분류에 의한 특허지도를 손쉽게 작성하는 데 그 목적이 있다.

현재 한국에는 WIPS라는 특허정보시스템에서 특허지도를 작성하는 소프트웨어 도구(tool)가 개발되었고 이와 유사한 기능을 가진 것으로 한국특허청에서 PIAS(Patent Information Analysis System)을 개발하였다. 이와 같이 특허정보를 소프트웨어적으로 처리하여 특허지도를 그리는 시스템을 모두 망라해서 이하에서는 “특허지도작성시스템”이라 한다.

현재 한국에 있는 이 시스템들을 살펴보면 (그림 1)과 같이 구성되어 있다. (그림 1)은 외부데이터베이스(A)에서 서지적사항, 초록, 청구범위, 발명의 상세설명을 다운로드 받아 특허지도작성시스템에 저장하는 것을 보여주는 시스템 구조도이다. 기존의 특허데이터베이스(KIPRIS, EPO, WIPS 등)를 연결하여 여기에서 필요정보(서지적사항, 청구범위, 초록 및 발명의 상세설명 등)를 자동으로 다운로드 받아 이를 이미 만들어 놓은 특허지도작성시스템의 데이터베이스(C)에 저장하도록 한다. 예를 들어 좀더 자세히 설명하면 다음과 같은 단계로 구분할 수 있다. 1단계로 USPTO(미국특허청) 특허데이터베이스에 들어가서 원하는 검색조건으로 검색한 후 해당 특허들을 찾아낸다. 예를 들어 검색어로 “QPSK”라



(그림 1) 특허지도작성시스템 구조도

고 하면 이 단어를 포함한 특허리스트가 웹사이트 화면에 나타난다. 2단계로 해당 리스트가 뜨면 이를 자동적으로 다운로드 받아 내부데이터베이스에 저장한다. 1단계 작업을 2단계 작업으로 처리함에 있어 연동장치가 필요하다. 연동장치가 필요한 이유는 여러 가지 있지만 첫째, (그림 1)에 보여준 4개의 데이터베이스를 보아도 접속하는 방법이 각 데이터베이스(A)마다 다르다. 둘째 특허데이터베이스(A)는 세계적으로 여러 종류의 데이터베이스가 있고 이를 표현하는 방법이 각각 다르다. 예를 들면 출원번호를 어느 데이터베이스는 1997-0012345로 표현하는데도 있고 또 다른 데이터베이스는 1997-12345라고 표기할 수 있기 때문에 이를 하나의 형태로 만들어 주는 작업이 필요하다. 이를 내부데이터베이스의 형태에 맞게끔 변환시켜야 한다. 이와 같은 작업을 연동

모듈에서 처리해준다. 하지만 방법에 따라 접속하여 데이터 다운로드는 연동모듈(20)에서 처리하고 데이터 변환장치는 특허지도작성시스템의 데이터베이스(C)에 저장한 후 서로 다른 형태의 데이터를 동일형태로 바꾸는 작업을 할 수도 있다.

특허지도작성시스템의 데이터베이스(C)에 저장된 특허는 일련의 작업을 거쳐 동일한 특허를 삭제하거나 대체시키는 작업을 통하여 필요한 특허만을 저장하게 된다. 즉 여러 데이터베이스(A)에서 받은 특허정보를 연동모듈을 통하여 다운로드 받아 데이터베이스(C)에 저장하는 방법은 널리 알려진 기술이다. 작업자가 원하는 만큼 데이터를 정리한 후 특허지도작성모듈(D)을 통하여 상기에서 설명한 여러 종류의 맵을 만들 수 있다.

특허지도작성시스템의 하나인 PIAS에서는 기술

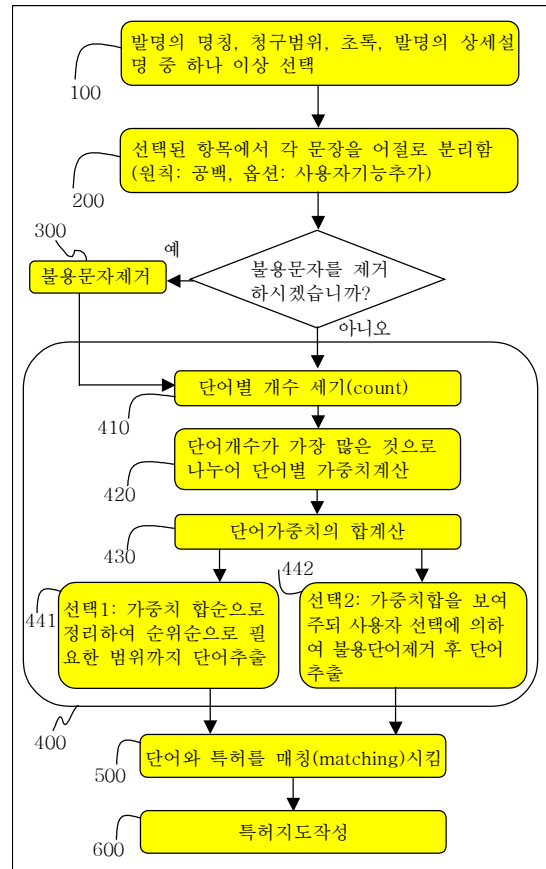
분류별 특허지도 작성을 위해 다음과 같은 방법을 사용하고 있다. 특허마다 기술별로 분류작업을 하려면 발명의 내용을 보지 않고 기술별로 분류하기는 어렵다. 그래서 각 특허에 대하여 초록 또는 명세서를 읽어보아 특허마다 자신이 분류한 기술분류에 할당하게 하는 방법을 취하고 있다. 예를 들어 PSK (Phase Shift Keying)라는 위상변조방법이 있는데 이것이 DPSK, QPSK, BPSK 등을 기술별로 보고자 할 때 명세서를 보고 각 특허에 대하여 일일이 이를 지정하여 분류하는 방법이다. 이렇게 기술분류한 다음 원하는 형태의 기술분류를 한 후 여러 가지 형태의 특허지도를 만들 수 있다. 지금까지 알려진 기술은 여기까지 이다.

하지만 이는 이 작업을 수행하는 자가 일일이 명세서를 보아야 하는 번거로움이 있고 이 때문에 특허지도의 활성화가 어려워 특허지도의 무용론까지 대두되고 있다. 하지만 이점을 본 아이디어를 통하여 쉽게 하면 특허지도작성이 수월해져 기술을 개발하는 사람에게 많은 도움을 줄 수 있을 것으로 기대된다.

본 아이디어는 상기의 종래 문제점을 개선하기 위하여 이를 자동화하는 방법을 제공하기 위한 것이다. 이 아이디어의 착안은 “중요한 기술적 단어(기술 용어)는 발명의 명칭, 초록, 발명의 상세한 설명에서 자주 언급된다”는 데 있다는 것이다. 이 아이디어는 반드시가 아니라 확률적으로 다분히 맞다는 것이다. 따라서 본 아이디어는 이를 이용하여 자동화하는 방식을 기본적으로 제시하고 이의 한계점을 보완하는 반자동화 하는 방법을 제시하고자 한다. 즉 일일이 초록이나 발명의 상세설명을 읽지 않고 단어를 추출하여 특허와 단어를 손쉽게 매칭시키고자 하는 데 본 논문의 존재이유가 있다.

실제 단어를 추출하는 것은 주로 기술적 단어를 추출하여 사용하는 게 가장 기본적인 방법기술적 단어가 아닌 특정한 목적의 단어를 중심으로 분류할 수 있으므로 표현의 정확성을 위해 “기술적 단어”라는 말보다는 “단어”라는 표현을 사용하도록 한다.

자동화를 구현하기 위한 실시 예로 첨부된 (그림



(그림 2) 특허지도작성 처리흐름도

2)를 이용하여 설명하면 6단계로 크게 구분할 수 있다. (그림 2)는 특허지도작성시스템에 저장된 데이터로부터 단어별 특허분석하여 특허지도를 그리기까지의 처리흐름도를 나타낸 것이다.

1단계는 다음과 같다(100). 본 발명에 사용된 특허지도작성시스템은 각종의 데이터베이스에서 가져온 서지적사항(이하 서지적사항이라 함은 초록을 포함하여 말한다)을 자체 데이터베이스((그림 1)의 C)에 저장한다. 이중에 단어를 포함하고 있는 것은 발명의 명칭, 초록, 청구항, 발명의 상세설명이다. 이 중에서 단어를 추출할 것의 4개 중 하나 이상을 선택한다. 가장 전형적으로 사용될 수 있는 것은 청구항이고 발명의 상세설명은 엄청나게 빠른 속도의 컴퓨터로는 가능하지만 느린 속도의 컴퓨터로 처리하기에는 부하가 많이 걸릴 수 있다. 그리고 대개 한 가

지 항목만을 선택하지만 필요에 따라서는 2개 이상의 항목을 단어추출대상으로 삼을 수 있다. 여기에서는 일실시 예로 청구항을 중심으로 설명하기로 한다. 따라서 본 아이디어는 청구항만을 기준으로 하는 게 아니고 본 방법을 다른 나머지 대상에도 적용시킬 수 있음을 분명히 한다. 즉 4개의 대상(발명의 명칭, 청구항, 초록, 발명의 상세설명) 중에서 어느 한 가지 이상의 대상을 선정하는 작업까지를 1단계로 한다.

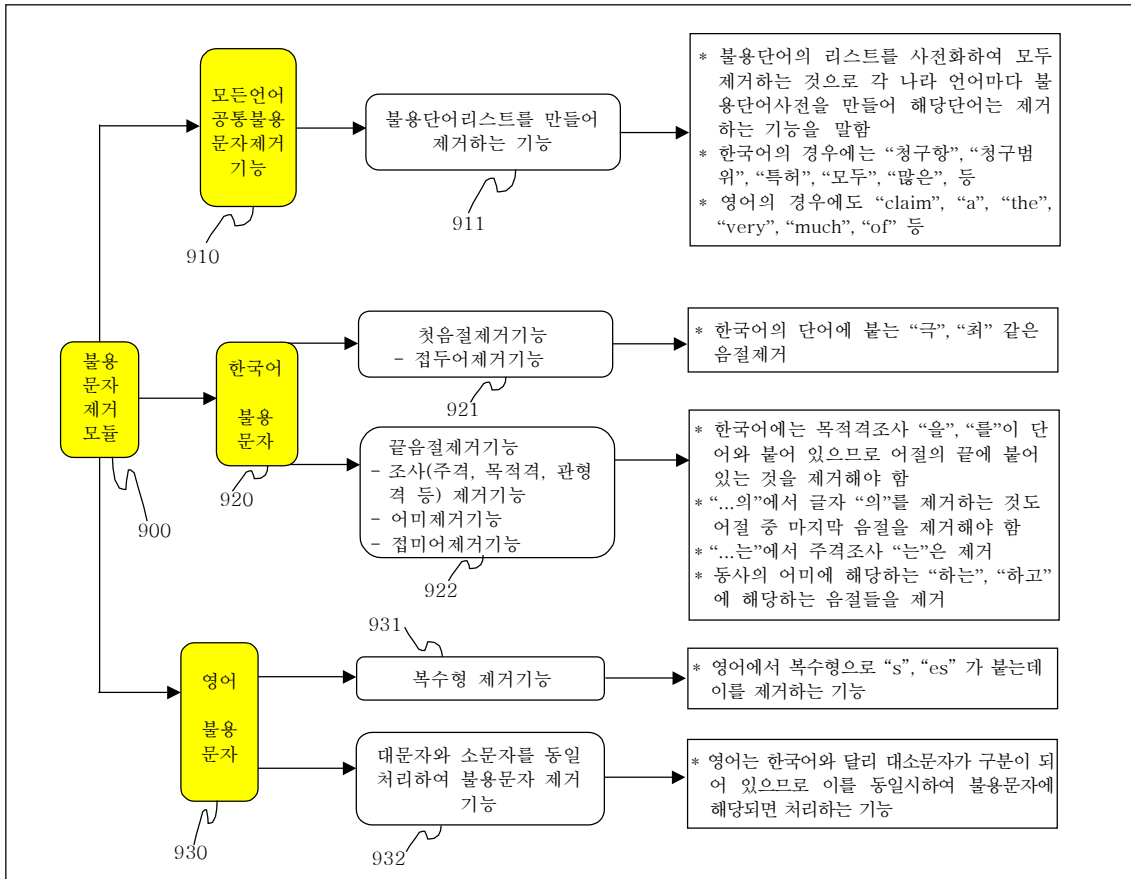
2단계는 다음과 같다(200). 특허지도작성시스템의 데이터베이스(그림 1)의 C)에 들어가 있는 청구항을 어절로 분절하는 작업을 말한다. 가장 전형적인 방법은 어절을 공백(space)으로 나누는 방법이다.

하지만 사용자에게 따라서 특수문자 등을 포함시킬 수 있다. 예를 들어 괄호, “/”, 콜론, 세미콜론 등이다. 만일 사용자가 이와 같은 문자를 어절로 분리한다고

지정하면 어절로 분리하는 기준이 될 수 있다.

3단계는 다음과 같다(300). 본 단계는 필수적 단계가 아니고 선택적 단계(option)이다. 3단계를 거치면 뒷단계에서 작업이 많이 편해지고 본 단계를 거치지 아니하면 뒷단계가 시간이 많이 소요되고 불필요한 단어를 거르는 작업이 필요할 수 있다. 본 단계를 넣은 이유는 어절로 분리되었기 때문에 여기에서 필요없는 음절은 제거하고 필요한 단어만을 추출해내야 하기 때문이다.

3단계 작업을 하기 위해 필요한 모듈이 불용문자 제거모듈이고 이 기능을 수행하는 기능은 (그림 3)에 잘 나타나 있다. 단어추출을 위하여 각 언어상의 특성으로 처리하는 방법이 다르다. 이를 크게 나누어 모든 언어영역에 해당하는 분야와 한국어와 영어를 중심으로 살펴본다.



(그림 3) 불용문자제거모듈의 세부기능

먼저 세계 모든 언어에서 적용되는 불용단어리스트 제거모듈을 보자(910). 단어를 추출할 때 일단 필요없는 단어리스트 또는 단어사전을 만들어 전형적으로 불필요한 단어를 제거하는 것이 좋다. 예를 들면 한국어의 경우를 보면 “잘”, “매우”, “그리고”, “또는” 등과 같은 단어이겠고, 영어의 경우는 “a”, “the”, “this”, “that”, “much”, “little” 등과 같은 단어가 있고 독일어의 경우도 “des”, “dem”, “den” 같은 단어, 프랑스어도 “La”, “De” 같은 단어 등 불필요한 단어는 어느 나라 언어이든 있기 마련이다.

한국어의 언어특성상 생기는 불용문자제거모듈을 보자(920). 한국어불용문자제거기능에는(920) 어절에서 불필요한 음절을 제거하기 위해 처음음절제거기능(921)과 끝음절제거기능(922)을 제거하는 두 가지 기능이 있다. 처음음절제거기능이란 글 어절의 처음에서 “극”, “최” 같은 단어를 제거하는 것을 말한다. 끝음절제거기능(922)은 한국어에서 특히 필요하다. 한국어는 일본어와 달리 띄어쓰기를 하면서도 조사가 영어처럼 분리되지 않고 단어 뒤에 붙는다. 대표적인 것이 목적격조사인 “을”, “를”이고, 관형격조사인 “의”, 주격조사인 “는”, “은”을 들 수 있다. 그 이외에도 동사의 어미에 해당되는 것이 있는데 “하는”, “하고”, “하여” 등의 동사어미에 해당하는 것으로 조사가 뒤에 붙어 있는 경우이다.

마지막으로 영어불용문자제거기능(930)을 살펴 보도록 하자. 영어불용문자제거기능(930)에는 크게 2가지 복수어미제거기능(931)과 대소문자동일처리 기능(932)이 있다. 복수어미제거기능(931)이란 영어에서 복수형에는 “S”, “ES”가 붙는데 이를 제거하는 기능을 말한다. 대소문자동일처리기능(932)이란 알파벳이 대소문자가 다른 단어로 인식되기 때문에 이를 동일화하여 불용단어에 해당되는 것은 제거하는 기능을 말한다. 지금까지 3단계를 거치게 되면 유용한 단어들을 추출할 수 있음을 보았다. 만약 이 기능이 없으면 가중치를 보여주되 불필요한 단어제거기능(422)을 사용해야 한다. 가능한 3단계 기능을 포함하는 특허지도작성시스템을 만드는 게 바람직하다.

4단계는 다음과 같다(400). 3단계까지 거쳐 지금까지 유용한 단어들을 추출하였다. 청구항에서 단어를 추출하여 조사해 보니 다음과 같이 나왔다고 하자. 첫번째 특허에서는 단어A가 20번, 단어B가 15번, 단어C가 10번, 단어D가 5번 나왔다고 하자. (410) 이때 최고 빈도수 단어는 단어A 이다. 단어A는 20번 나왔으므로 단어빈도수를 최고 높은 값 20으로 나누면 단어A가 1, 단어B가 0.75, 단어C가 0.5, 단어D가 0.25가 된다(420). 두번째 특허에서 단어빈도수를 조사해보자. 두번째 특허에서는 단어A가 5번, 단어B가 10번, 단어C가 0번, 단어D가 5번, 단어F가 5번 나왔다고 하자(410). 이때 다시 단어 최고 빈도수 10으로 나누어 가중치를 구하면 단어A가 0.5, 단어B가 1, 단어C가 0, 단어D가 0.5, 단어F가 0.5가 된다(420). 상기와 같은 방법을 각 특허들에 적용하면 각 단어별 가중치가 만들어 진다. 이렇게 만들어진 단어별 가중치를 구하여 이를 합하면(430) 중요단어 순으로 정렬할 수 있다. 예를 들어 50개의 특허를 분석한다고 할 때 가중치의 합이 단어A가 40, 단어B가 20, 단어C가 7, 단어D가 20, 단어F가 10이 나왔다고 하자. 이를 처리하는 방법은 사용자의 의도에 따라 두 가지 방법을 택할 수 있다. 하나는 가중치순으로 정리하여 필요한 범위 내까지의 순위만 채택하는 방법(441)과 가중치의 합으로 보여주되 불필요한 단어는 제거하는 기능(442) 두 가지가 있다. 먼저 첫번째 방법(441)을 살펴보자. 위의 경우 가중치 순서대로 하면 단어A, 단어B, 단어D, 단어F, 단어C 순이 된다. 이때 세번째까지만 필요하면 단어A, 단어B, 단어D만 취하고 나머지 데이터는 버리도록 한다. 이렇게 하여 단어A, 단어B, 단어D를 기술분류기준으로 채택하면 50개의 특허는 각 단어별 분류가 가능하게 된다. 두번째 방법인 불필요한 단어제거기능을 취하는 방법(442)을 살펴보자. 위의 경우 가중치 순서대로 하면 단어A, 단어B, 단어D, 단어F, 단어C 순이 된다. 이중 가중치가 아무리 높아도 불필요한 단어라고 생각되면 제거할 수 있도록 하는 기능이다. 만일 단어A, 단어D, 단어F가 불필요한 단어라고 생각되면 단어B, 단

어F, 단어C만을 채택한다.

5단계는 다음과 같다(500). 특허와 추출된 단어를 매핑시키는 기능을 5단계에서 수행한다. 특허를 원칙적으로 1:1로 매핑시키는 게 일반적이지만 상황에 따라서는 1:N 복수 매핑도 가능하다. 예를 들면 한 특허가 단어A, 단어B 2개 이상을 포함하는 경우인데 이를 어떻게 처리할 것인가가 문제가 된다. 이때에도 이는 기능선택의 문제로서 단어A가 많은지 단어B가 많은지를 비교해서 처리할 수도 있고 (1:1로 매핑) 동시에 2곳에 포함시킬 수 있다(1:N 복수 매핑). 이는 사용자의 의사에 따라 세팅함으로써 처리 가능하다.

마지막 6단계는 다음과 같다(600). 일단 매핑작업이 끝났으면 기존의 특허지도를 그리듯 하면 된다. 특허지도작성시스템의 데이터베이스(그림 1)의 C)는 출원일, 출원번호, 국가, 등록일, 등록번호, 대표 발명자 등 여러 가지 서지적사항이 있다. 지금까지 작업은 기술분류에 사용되는 단어를 추출하여 데이터베이스의 한 필드값을 만든 일련의 작업이다. 따라서 기술분야별 연도별 분석, 기술분야별 국가별 분석, 출원인별 기술별 분석 등 다양한 형태의 특허 맵을 만들 수 있다.

지금까지 6단계를 거쳐 특허지도를 자유자재로 그릴 수 있지만 이 기능을 확장할 수 있다. 이는 최초의 단계만 분류한 것이다. 비유를 들어 설명하면 처음에 위상변조에 해당하는 특허를 모아다가 PSK 변조방식에서 “BPSK, QPSK, DPSK” 등으로 구분

한 경우이다. 하지만 “QPSK”를 다시 분류할 수도 있다. 만일 “QPSK”에서 “변조”와 “복조” 기술로 분류하고 싶을 수 있다. 이런 때에 현재 데이터베이스에 매핑되어 입력된 데이터를 가지고 재분류 과정을 거칠 수 있다. 이런 일련의 반복적(recursive) 작업을 통하여 N번까지 기술분류가 가능하다.

VII. 결론

본 논문을 통하여 PM의 경향을 살펴보았고 작성 방법 등을 알아보았다. 그리고 이상에서 설명한 바와 같이 본 고에서는 기술분류표(예: IPC 분류)에 의하지 아니하고 특허에서 중요단어를 추출하여 자동으로 분류하는 것으로 훨씬 사용자의 요구에 부응할 수 있으며 손쉽게 기술흐름을 파악할 수 있는 PM 작성툴에 대하여 살펴보았다. 본 고에서 제시한 아이디어는 PM을 손쉽게 작성하는 방법으로 특허청에서 수용여부를 기대하며 후에는 이보다 더 나은 시스템들이 나올 것을 기대해 마지 않는다.

참 고 문 헌

- [1] <http://www.patentmap.or.kr/>
- [2] 특허정보분석시스템 매뉴얼 ver2, 특허청, 2001. 11.
- [3] 특허 맵 개념 및 활용, 특허청, 2001. 3.
- [4] 인터넷특허정보검색, 특허청