

# 텍스트마이닝을 활용한 특허관리 방법론

글 \_ 윤병운 · 서울대 산업공학과 기술경영연구실 · postman3@cybernet.snu.ac.kr

## 1. 새로운 특허관리 방법론의 필요성

특허는 기술혁신의 다양한 현상을 분석하는데 거의 유일한 정량적 자료로 받아들여지고 있다. 이것은 특허를 출원하고 등록하는 과정이 제도적으로 규정되어 있어 기술 혁신 활동에 대해 신뢰성 높은 자료를 양산하며, 정보통신 기술의 발달로 방대한 양의 자료를 손쉽게 습득할 수 있기 때문이다. 또한 최근 들어 비즈니스 모델에 대한 특허가 세계 각국의 특허청에서 인정됨에 따라 과거의 기술 중심의 혁신에서 영역을 확대하여 서비스 혁신에 대한 분석도 가능해졌다. 따라서 특허의 분석은 기술개발활동 뿐만 아니라 사업 개발 활동에 이르기까지 활용가치가 확장되었다고 할 수 있다.

특히 특허 분쟁이 첨예화되어 소송 결과가 기업의 존폐로까지 연결되거나, 크로스 라이선스(cross license)를 통해 기술 협력하는 사례가 빈번해지면서, 기업의 특허를 관리하고 분석하는 행위는 기술경영에서 매우 핵심적인 업무가 되어가고 있다.

이러한 업무를 위한 분석을 가능하게 하기 위해서는 풍부한 특허 정보를 가공하고 활용하는 방법의 제시가 필수적이다. 그러나 지금까지 특허 관련 연구들은 앞서 제시한 요구를 충족할 만한 방법론을 제공하기에 매우 미흡했다.

첫째, 연구의 데이터로 활용하는 정보가 출원연도, 출원국, 출원개수 등과 같은 서지적 정보에 국한되어 특허 문

서가 서술하고 있는 기술적, 권리적 내용에는 초점을 맞추지 못했다.

둘째, 분석 방법론 차원에서 회귀분석, 상관분석 등과 같은 다변량 통계 분석(multi-variate statistical analysis)의 활용에 집중되어 있어 방대하고 유용한 정보를 제한적으로 활용, 분석하고 있다.

셋째, 분석의 내용 측면에서는 대부분 거시적 차원의 국가, 산업의 기술혁신 활동 분석에 초점을 맞추고 있어 공백기술의 발견이나 기술저축 탐색, 신기술 개발과 같은 기업 및 기술 차원의 분석은 거의 찾아볼 수가 없다.

따라서 새로운 특허 분석 방법론은 특허 문서의 풍부한 기술적, 권리적 정보를 지도나 네트워크, 매트릭스와 같은 다양한 형태를 통해 시각화하고, 기존의 특허분석이 제공하지 못하는 기업의 기술개발 활동 영역을 지원할 수 있어야 할 것이다.

이를 위하여 새로운 방법론은 텍스트마이닝(text mining), 네트워크 분석(network analysis), 뉴럴 네트워크(neural network) 등과 같은 데이터마이닝 방법이나 사회과학적, 경영과학적 방법론들을 이용한 학제간(multi-disciplinary) 연구의 성격이 매우 짙다고 할 수 있다.

## 2. 텍스트마이닝

텍스트마이닝은 구조화되어 있지 않은 문서로부터 자동적으로 정보를 추출하고 패턴을 분석하는 것이다[Dixon, 1997]. 이 방법론은 일반적으로 인터넷 검색 엔진이나 데이터베이스 관리 기술과 혼동되는 경우가 많지만 근본적인 특성상, 텍스트마이닝은 이들과 구분되어야 한다.

검색의 경우, 사용자는 이미 알려져 있고 다른 사람에 의해 작성된 문서를 찾아내는 것에 초점을 맞추지만, 텍스트마이닝의 궁극적 목적은 알려져 있지 않은 정보나 패턴을 도출해내는 것에 있다[Feldman and Dagan, 1995]. 즉, 방대한 양의 데이터들로부터 유용한 정보를 이끌어내는 데이터마이닝과 유사하게 텍스트마이닝은 구조화되지 않은 문서로부터 의미 있는 정보를 도출한다[Fayyad et al., 1996].

텍스트마이닝의 연구 영역은 몇 가지로 요약될 수 있다. 우선, 사전에 정의된 분류로 문서들을 구분하는 것과 관련된 연구들이다. 이들은 k-nearest neighbor(k-NN) 기법을 텍스트 분류(text classification)에 적용하거나[Cohen and Hirsh, 1998; Yang, 1994], 인공지능의 일종인 machine learning을 이용하여 문서들을 자동적으로 분류하는 연구가 대표적이다[Sebastiani, 2002].

두 번째 영역은 문서 내용의 유사성을 토대로 문서들을 클러스터링(clustering)하는 것이다. K-means나 계층적 클러스터링 알고리즘과 같이 통계에서 활용되는 다양한 기법이 문서를 그룹핑하는 목적으로 사용되었으며

[Voorhees, 1986; Dhillon and Modha, 1999], 데이터 마이닝 기법의 일종인 SOM(Self-Organizing Map)이 적용되기도 하였다[Kohonen, 1995].

세 번째 영역은 자연 언어로 구성된 텍스트를 구조화하는 연구들로서, 대부분의 연구들은 machine learning의 수리적 모델을 활용하여 텍스트의 내용을 키워드를 통해 구조적으로 표현하였고[Freitag, 1998; Hsu and Chang, 1999], 웹페이지 문서의 HTML에서 구조화된 데이터를 추출하는 연구도 인터넷 사용이 일반화된 환경에서 활발히 이루어지고 있다[Hammer et al., 1997]. 네 번째 연구 주제는 텍스트로부터 의미 있는 정보나 지식을 추출하는 것이다.

이러한 연구들은 문서의 주제를 찾아내거나[Clifton, 2003], 전체 문서의 내용을 자동적으로 요약하는 기능을 제공할 수 있다[Mani and Maybury, 1999]. 마지막 연구 영역은 문서들의 관계를 시각화하여 분석자나 사용자가 관련 문서들의 구조를 용이하게 파악하는 데 초점을 둔다.

문서의 내용과 문서 간 관계를 도형으로 표현하거나[Rohrer et al., 1998], 문서에 포함되어 있는 키워드들을 시각화함으로써 연구 주제들의 프로파일을 정리하고[Porter et al., 2002], 기술의 발전 경향을 분석하기도 하였다[Watts and Porter, 1997].

## 3. 텍스트마이닝 기반 특허관리 방법론의 제안

텍스트마이닝을 활용한 새로운 특허관리 방법론으로서 키워드 기반 특허 지도(patent map), 특허 네트워크(patent network), 형태 분석(morphology analysis) 등 3가지 형태를 제시하고자 한다. 이 방법론들은 기술 관리 및 신기술 개발을 위해 특허문서에서 텍스트마이닝을 통해 키워드를 추출하고 이들을 활용하여 모든 특허문

서를 키워드 벡터화하여 구조화한다. 즉, 추출된 키워드는 벡터의 필드(field)가 되어 특허 문서들에서 각 키워드가 발생하는 빈도를 토대로 해당 필드가 채워지며, 모든 특허는 키워드 벡터화되고 이들은 궁극적으로 지도나 네트워크 등으로 시각화된다. 따라서 자연어로 처리되어 있는 특허문서를 구조화시키기 위한 텍스트마이닝의 적용

과정은 3가지 모듈에 공통적으로 포함되지만 각 모듈의 목표나 대표적인 활용 방법론은 상이하다. [표 1]은 3가

지 형태의 키워드 기반 특허 분석의 형태와 목표, 대표적 방법론을 정리한 것이다.

구분	특허 지도	특허 네트워크	Keyword-based MA
형태	특허를 평면에 매핑시켜 지도 작성	특허간의 관계를 네트워크로 작성	특허의 형태를 매트릭스로 표현
목표	공백기술의 영역 탐색, 저축 특허의 검색, 특허 포트폴리오의 작성	핵심 특허, 기술수명주기 분석, 클러스터의 작성	신기술의 형태 도출, 경쟁 기업의 기술경쟁력 분석
대표적 방법론	SOM	Network Analysis	Morphology Analysis

〈표 1〉 각 모듈의 형태와 목표

**(1) 특허 지도**

특허 지도는 특허의 서지적, 기술적, 권리적 사항을 명확하게 하고 이를 구체적인 목적에 따라 분석하고 해석하여 조합함으로써 도출한 결과를 한 눈에 파악할 수 있도록 그래프나 그림으로 표시한 것이다.

이것은 1960년대 일본에서 처음 제기된 이후로 실무적으로 활발하게 이용되고 있으며 한국 특허청에서도 2000년부터 매년 특허지도도를 개발하여 발표하고 있다.

그러나 기존 특허 지도는 다음과 같은 한계점을 가지고 있다. 우선 개발 방법에 대한 문제로서 관련 기술에 대한 전문가들의 장시간 협업에 의존하기 때문에 비용과 시간의 문제가 발생한다는 점이다.

또한 형태와 개발 도구 측면에서는 대부분의 기존 특허 지도가 테이블이나 2차원 그래프를 활용하고 평균이나 분산과 같은 매우 기초적인 통계량을 이용하는데 그치기 때문에 분석의 내용이 추세 분석이나 동향 파악 정도에 국한될 수밖에 없다는 점이다.

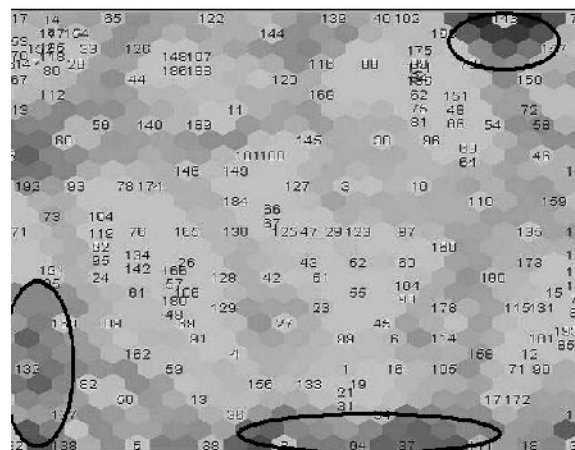
특히 특허 지도 작성에 활용되는 자료 측면에서 특허 문서에 포함된 기술적 내용을 적극적으로 이용하지 못한다는 문제가 있다. 따라서 키워드 기반 특허 지도는 기존 특허 지도에서는 분석하기 어려운 공백기술 영역의 도출이나 기술저축 여부 탐색, 포트폴리오 작성 등을 특허 문서가 포함하고 있는 키워드를 이용하고 최근에 활발히 이용되는 시각화 방법인 SOM을 이용하여 개발하였다[Yoon et al., 2002].

특허지도의 형태는 다양하게 제시될 수 있으나 대표적으로 기술의 유망한 영역을 도출할 수 있는 기술 공백 지도, 새로운 기술의 개발 가능성 탐색을 위한 기술 저축 지도,

기술의 유사성을 토대로 특허를 군집화하기 위한 기술 포트폴리오 지도 등이 있다. 미국 특허청에 출원된 DRAM 관련 특허를 토대로 기술 공백 지도의 개발 결과를 예시하면 다음 [그림 1]과 같다.

위에서 제시한 특허지도는 192개의 DRAM 관련 특허의 문서로부터 키워드를 추출하고 이를 토대로 키워드 벡터로 전환하여 특허의 유사성을 토대로 2차원 평면에 위치시켰다. 이 지도에서 공백 영역은 특허들의 분포와 노드들의 음영을 기반으로 평가될 수 있을 것이다.

[그림 1]에 타원으로 표시된 영역은 특허들이 이 영역에 드물게 포함되어 있고, 노드들의 음영이 짙게 표시되어 노드들 간 거리는 매우 먼 것으로 분석되어 있기 때문에 공백으로 판단될 수 있다. 공백의 의미는 전문가에 의해 상세하게 분석되어 개발 가치의 여부를 결정할 수 있어야 한다.



〈그림 1〉 DRAM 관련 특허의 기술 공백 지도

**(2) 특허 네트워크**

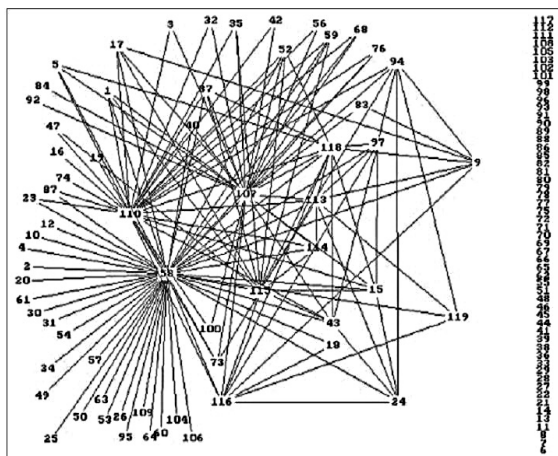
특허의 연관관계 분석은 공통 분류(co-classification) 분석, 공통 인용(co-citation) 분석, 공통 단어(co-

word) 분석으로 나누어 볼 수 있다. 공통 분류나 공통 인 용 분석도 빈번하게 활용되는 방법이지만 최근에는 기술의 본질적 내용을 토대로 특허 간 연관관계를 도출한다는 측면에서 공통 단어 분석을 활발히 이용하는 추세이다.

키워드 기반 특허 네트워크는 특허 간 연관관계를 키워드를 통해 규명하는 방법으로서 각 특허들이 가지고 있는 키워드들의 빈도수를 계산하고 특허 간 유사성의 정도를 산출한다. 즉, 특허 간 키워드의 유사성 측면에서 특허들의 연관관계를 네트워크로 시각화하며, 특허를 노드(node)로, 특허간의 키워드 유사성 정도를 링크(link)로 간주하게 된다[Yoon and Park, 2004].

[그림 2]는 119개의 WDM 기술을 활용하여 특허 네트워크를 시각화한 예이다. 특허 네트워크를 통해 개별 특허들의 연관관계를 쉽게 파악할 수 있으며, 네트워크의 포함 여부를 분석할 수 있다.

네트워크 분석에서는 밀도(density), 중심성(centrality) 등과 같은 지표들을 통해 정량적 분석을 제공한다. 특허 네트워크에서도 정량 지표로서 기술 중심성 지수, 기술 수명 지수, 기술 키워드 포트폴리오 등과 같은 지수를 제안하였다. 이들은 특허 네트워크 상에서 기술의 역할이나 기술 수명과 같은 기술적 특성을 분석하는 기준으로 활용될 수 있다.



〈그림 2〉 WDM 기술의 특허 네트워크

### (3) 형태 분석

특허 지도와 특허 네트워크는 특허의 시각화를 통해 기술 관리를 하기 위한 방법론이지만 구체적인 기술 개발의 형태는 제시할 수 없다. 키워드 기반 형태 분석은 구체적이고 실질

적인 신기술 개발 기회를 탐색하기 위한 방법론이다. 형태 분석의 기본 개념은 특정 시스템이 몇 개의 차원(dimension)들로 분리될 수 있고, 이 차원들을 통해 전체 시스템이 가능한 포괄적이고 세부적으로 설명될 수 있다는 것을 바탕으로 한다[Wissema, 1976].

기본적으로 시스템은 많은 하위 시스템으로 구성되어 있고, 이들의 각 형태(shape)는 매우 다양하다. 형태 분석은 시스템을 설명하는 각 차원들이 취할 수 있는 다양한 형태를 정의하고 이 형태를 조합함에 의해서 시스템이 채택할 수 있는 모든 가능한 대안들을 조사하며, 기존 시스템의 형태를 확인함으로써 발현되지 않은 시스템 형태의 조합에 대해 신기술이나 신제품의 개발을 제안하는 방법이다. [표 2]는 형태 행렬(morphology matrix)의 예시이다.

Process	Energy	Structure	Function	Material
Addition	Chemical	Quadrat	Drilling	Stone
Removal	Heat	Triangle	Grilling	Iron
Grouping	Biological	Circle	Turning	Gas
		Linear		Paper

〈표 2〉 형태 행렬의 예시

특정 기술을 표현하는 데 있어 process, energy, structure, function, material 등 5개의 차원이 필요하고 각각은 3, 4개의 형태를 취하게 된다. 회색으로 표시된 셀은 특정 기술이 각 차원에서 취하게 되는 형태를 표시한 것이다. 관련된 모든 기술의 형태 조합을 확인할 수 있다면 미개발 형태(configuration)를 도출해낼 수 있을 것이며 이것은 신기술 개발의 잠재적 대상이 될 수 있을 것이다[Yoon and Park, 2005].

키워드 기반 형태 분석은 이러한 형태 확인의 과정을 각 특허가 보유하고 있는 키워드를 통해 정량적, 체계적으로 지원하는 방법론이다. 가령 특허 문서가 'addition'이나 'grouping'보다 'removal'과 관련된 키워드를 더 많이 포함하고 있다면 이 특허의 'process'라는 차원은 'removal'로 표시된다. 다른 차원에서의 형태들도 이와 같은 방법으로 결정되며 수집된 특허의 형태(configuration)를 모두 정리한 후에 존재하지 않는 조합(unoccupied combination)을 도출할 수 있게 된다.

결과적으로 키워드 기반 형태 분석은 텍스트마이닝이라는

도구와 특히 정보라는 정량적인 자료를 채택함으로써 기술의 미개척 분야(terra incognita)의 가능한 집합을 제공하게 된다. 이러한 조합들 중에서 우선순위(priority)를 정해줄 수 있다면 연구개발 기획이나 프로젝트 관리 업무를 지원할 수 있을 것이다. [표 3]은 TFT-LCD 기술에서 광시야각 관련 기술에서의 형태 목록을 정리한 것이다.

사전에 정의된 형태 구조(morphology structure)에서 A에서 G까지의 7개의 차원들은 각각 3, 4개의 형태를 띠고 있으며 각 특허들의 형태는 자신이 가지고 있는 키워드를 통해 결정된다. [표 3]에서 'A3'는 첫 번째 차원에서 3번째 형태를 취한다는 표시이며 첫 번째 형태(morphology)를 취하는 특허는 모두 4개가 존재한다는 것을 나타내고 있다. 이와 같이 관련된 모든 특허의 형태

를 확인한 후에, 미개발된 기술의 형태를 정리하고 기술 가치평가 등의 기준을 토대로 이들의 우선순위를 정하게 된다. 이런 과정을 통해 신기술 개발의 형태를 분석, 지원해줄 수 있을 것이다.

Morphology	Patent Number	Total Number
A3-B1-C1-D2-E1-F2-G1	5,594,570, 5,652,634, 5,689,322, 5,855,968	4
A2-B1-C1-D3-E3-F2-G1	6,330,108, 5,847,688, 6,399,165, 6,519,018	4
A2-B1-C1-D3-E3-F2-G2	6,529,256, 6,583,839, 6,587,841, 6,707,523	4
A3-B1-C2-D3-E1-F2-G1	6,417,905, 6,433,850, 6,191,836, 6,633,355	4
A2-B2-C2-D2-E3-F2-G1	6,335,776, 6,449,025, 6,665,035	3

〈표 3〉 광시야각 기술의 형태 목록

#### 4. 결론 및 추후연구 과제

21세기 세계의 경제는 지식기반경제에 직면하고 있으며 기업은 경쟁력 확보를 위해 지식경영 활동을 장려하고 있고, 정부는 지식집약 산업을 적극적으로 지원하고 있다.

그러나 이들의 노력은 지식에 관련된 현상을 연구하거나 이미 개발된 지식을 관리하는 데 초점을 맞추고 있고, 지식 창출의 구체적이고 체계적인 과정 및 방법론을 제시하는 데는 매우 미흡하다.

더욱이 지식경영 시스템들은 효율성, 효과성 측면에서 낮은 평가를 받으며, 대부분의 지식경영 도입은 실패로 판명되고 있다. 이러한 상황은 신기술 개발을 포함한 기술 관리의 측면에서도 유사하게 적용된다. 많은 연구들과 시스템이 개발, 도입, 실행되고 있지만, 기술의 현상을 분석하는 데 그치는 학문적인 것이거나 신기술 개발이나 기술 관리에 있어 가치 있는 결과를 출력하지 못하는 시스템들이 대부분이다.

텍스트마이닝을 이용한 새로운 특허 관리 기법들은 기술 지식을 관리하고, 잠재된 의미를 찾아냄으로써 새로운 기술지식을 창출하고 전략적으로 활용하는 데 목적이 있다.

특허 문서는 이러한 목적을 달성하기 위한 기초 데이터로서 선정되었고, 이를 실질적인 분석에 이용하기 위한 전 처리 방법론으로서 텍스트마이닝을 적용하였다. 또한 핵심적인 방법론의 내용으로서 특허 지도, 특허 네트워크, 형태 분석

을 제안하여 특허를 2차원 평면이나 네트워크로 시각화하고 실질적인 신기술 개발의 형태를 추천하는 역할을 한다. 그러나 제안된 방법론들은 몇 가지의 한계점을 지니고 있다. 우선, 데이터 수집의 과정에서 해당 기술과 관련된 특허를 검색하는 것은 전적으로 특허청에서 제공하는 데이터베이스에 대한 관련 키워드 검색의 방법에 의존한다. 그러므로 방법론에서 활용하는 문서 형태의 데이터를 수집하는 과정 및 방법에 대한 명확한 고려가 있어야 한다.

둘째, 키워드를 기반으로 기술을 분석하고자 하는 연구는 문서로부터 1차적으로 도출된 키워드들 중에서 특허 문서를 대표하는 것이 아닌 특허를 대표하는 키워드를 선별하는 것이 핵심이다. 따라서 기술의 온톨로지(ontology)를 키워드 기반 특허 분석에 접목시키는 방법론에 대한 폭넓은 연구가 수행되어야 할 것이다.

셋째, 특허 지도나 특허 네트워크의 시각화의 경우, 지도의 종류나 네트워크 지표의 활용이 매우 제한적일 수 있다. 지도의 경우, 공백 기술, 기술 저축, 기술 포트폴리오를 제시하였고, 네트워크 지표의 경우에는 기술 중심성, 기술 수명이 제안되었으나 기술을 설명하고, 전략적인 판단을 목적으로 하는 분석에서는 다양한 지도 및 지표가 제안되어야 할 것이다.

넷째, 기술 관리에서는 컴퓨터에 의해 자동화될 수 있는 부분과 전문가가 담당할 수 있는 부분이 혼재해 있다. 키워드 기

반 특히 분석에서는 키워드의 선택, 기술 사전의 적용 등에서 전문가의 역할을 강조했지만, 이에 대한 구체적인 가이드라인(guideline)을 제시하지 못했다.

따라서 향후 연구에서는 각 방법론들을 적용하는 과정에서 전문가의 작업과 자동화 작업 등을 명확히 분리하는 논리, 내용 및 과정을 제시할 수 있어야 한다. 다섯째, 기술은 도메인에 따라 매우 다른 특성을 보이게 된다. 제품 기술, 공정 기술의 구분뿐만 아니라, 본질적인 기술들의 특성은 매우 다를 것이다. 제안된 방법론들은 기술에 대한 특성의 고려 없이 일반적인 경우의 적용에 초점을 맞춘 보편적 모델(generic model)이다. 따라서 기술의 상이한 특성들을 방법론에 포함

시켜 맞춤(customization) 분석을 가능하게 하는 연구가 시도될 수 있다.

키워드 기반 특허 분석은 기술 관리를 시스템적으로 실행할 수 있는 방법론에 관련된 연구로서, 기술 개발자뿐만 아니라, 보유 기술의 관리자, 기술 개발의 기획자, 중장기 기술 발전 전략을 구상하는 경영자, 그리고 국가 전체의 기술 정책 입안의 책임자들에게 기술 발전의 방향이나 기술 개발의 전략 수립, 구체적 신기술의 형태 등을 파악할 수 있는 중요한 정보를 제공해 줄 수 있다. 향후 텍스트마이닝을 이용한 특허 분석은 그 활용 영역을 확장하고 방법론적으로 보완되어 실무에서 폭넓게 활용될 수 있을 것이다.

## 참고문헌

- Clifton, C., "TopCat: Data Mining for Topic Identification in a Text Corpus", *IEEE Transactions on Knowledge and Data Engineering*, 16(8), 949-964, 2004.
- Cohen, W.W. and Hirsh, H., "Joints that generalize: text classification using WHIRL", In proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining, New York, 1998.
- Dixon, M., "An overview of document mining technology", unpublished paper, 1997.
- Elzen, B., Enserink, B., and Smit, W.A., "Socio-technical networks: how a technology studies approach may help to solve problems related to technical change". *Social Studies of Science*, 26 (1), 95-141, 1996.
- Fayyad, U., "From data mining to knowledge discovery: An overview", In Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R(Eds), *Advances in Knowledge Discovery and Data Mining*, MIT Press: Cambridge, Mass., 1996.
- Feldman, R. and Dagan, I., "Knowledge discovery in textual databases(KDT)", In proceedings of the first International Conference on Knowledge Discovery and Data Mining (KDD-95), Montreal, Canada, 1995.
- Freitag, D., "Information extraction from HTML: Application of a general machine learning approach". In proceedings of the 15th Conference on Artificial Intelligence (AAAI-98). Menlo Park, CA, 1998.
- Hammer, H., Garcia-Molina, J., Cho, R., Aranha, A., and Crespo, V., "Extracting semistructured information from the Web", In proceedings of the Workshop on Management of Semistructured Data (PODS/SIGMOD'97), Tucson, Arizona, 1997.
- Hsu, C.N. and Chang, C.C., "Finite-State Transducers for semi-structured text mining", In proceedings of International Joint Conference on Artificial Intelligence (IJCAI) Workshop on Text Mining, Stockholm, 1999.
- Kohonen, T., *Self-organizing maps*, Springer: Berlin, 1995.
- Mani, I. and Maybury, M.T., *Advances in Automatic Text Summarization*, The MIT Press: Cambridge, 1999.
- Porter, A.L., Kongthon, A., and Lu, J.C., "Research profiling: improving the literature review", *Scientometrics*, 53(3), 351-370, 2002.
- Rohrer, R.M., Ebert, D.S., and Sibert, J.L., "The shape of Shakespeare: visualizing text using implicit surfaces", In proceedings of the 1998 IEEE Symposium on Information Visualization, North Carolina, 1998.
- Sebastiani, F., "Machine Learning in Automated Text Categorization", *ACM Computing Surveys*, 34(1), 1-47, 2002.
- Voorhees, E.M., "The effectiveness and efficiency of agglomerative hierarchic clustering in document retrieval", PhD thesis, Cornell University, 1986.
- Watts, R. J. and Porter, A. L., "Innovation forecasting", *Technological Forecasting Social Change*, 56, 25-47, 1997.
- Wissema, J. G., "Morphological analysis: its application to a company TF investigation", *Futures*, 8(2), 146-153, 1976.
- Yoon, B. U. and Park, Y. T., "A text-mining-based patent network: Analytical tool for high-technology trend". *Journal of High Technology Management Research*, 15, 37-50, 2004.
- Yoon, B. U. and Park, Y. T., "A systemic approach for identifying technology opportunities: keyword-based morphology analysis". *Technological Forecasting and Social Change*, 72, 145-160, 2005.
- Yoon, B. U., Yoon, C. B., and Park, Y. T., "On the development and application of a self-organizing feature map-based patent map", *R&D Management*, 34(4), 291-300, 2002.