

과학기술 정보분석 시스템 동향 및 향후 발전 방향

배상진*·여운동**·이상필***

I. 서론

과학기술 정보분석은 연구결과에 따르는 방대한 양의 과학기술 정보를 조사, 검토하여 연구 개발에 필요한 목적에 활용할 수 있도록 하는 과학기술 연구개발 활동의 한 부분으로, 유용한 과학기술정보의 스크리닝과 각종 분석도구를 이용한 과학기술정보의 고부가가치 가공을 통하여 수요자(연구개발자)의 R&D 활동에 활용토록하여 고부가가치의 지식을 창출하는 역할을 한다.

21세기에 들어서면서 OECD 국가들의 경제는 급속히 지식과 정보에 기반을 둔 경제체제로 바뀌고 있다. 이러한 창조적 지식은 생산성과 경제성장의 원동력으로 인식되고 있으며 경제적 성취에 있어서 정보, 기술 및 학습의 새로운 역할에 대해 관심이 집중되고 있다. 미국 과학재단(NSF)의 조사에 따르면 첨단 분야의 연구 개발자들이 정보를 조사하고, 기술 현황을 분석하는데 소요되는 시간과 비용이 전체 연구 활동의 약 40%에 이르고 있다고 한다. 정보의 홍수 시대에 살고 있지만 필요로 하는 정보를 조사하고 기술동향과 현황을 분석하는데 아직 너무 많은 시간과 비용을 투입하고 있는 것이다. 이에 따라 정보를 효율적으로 활용하는 방법, 정보를 통해서 임무에 적합한 지식을 찾아내는 정보분석이 점점 중요해지고 있는 것이다.

이러한 상황에 비추어 볼 때, 과학기술 정보분석의 목표는 분명한 활용목적 가지고 기술과 산업 등의 지식의 흐름을 파악하고 예측하는데 있다. 그리고, 정보화 사회에서 가장 필요로 하는 핵심 지식은 과학기술 활동에서 문제제기를 하는데 있다. 따라서 과학기술 정보분석은 정보화 사회의 연구개발 활동과 산업활동에 핵심이 되는 '문제제기'를 만들어내는 중요한 영역이며, 이러한 정보분석을 효과적으로 수행하는 수단을 과학기술 정보분석 시스템이라 한다.

1. 문제제기

과학기술관련 연구활동에 필요한 지식자원은 논문, 특허 등의 문헌정보 외에도 연구자, 연구기관, 연구시설, 커뮤니티정보, 산업시장정보 등 다양한 정보가 필요하다. 그런데 과학기술관련 연구활동에 필요한 이러한 과학기술 정보자원이 인터넷 이전에는 주로 공개되는 논문, 특허를 중심으로 조사되었지만, 인터넷의 발달로 개인 연구자의 능력과 역량으로 수집되었던 정보들이 이제는 쉽게 접근할 수 있게 되었다. 그런데 이러한 다양한 정보의 대부분이 온라인을 통해서 공개되면서 가용 가능한 정보는 점점 많아지고 있다.

이렇게 많은 과학기술정보를 어떻게 과학기술관련 연구활동에 효율적으로 활용할 것인가 하는 문제는 연구개발자의 고민거리가 아닐 수 없다. 이와같이, 연구활동에 있어서 과학기술정보의 조사 및 분석은 매우 중요하다 하겠다.

* 배상진, 한국과학기술정보연구원 선임연구원, 02-3299-6032, sjbae@kisti.re.kr

** 여운동, 한국과학기술정보연구원 연구원, 02-3299-6017, wdyeo@kisti.re.kr

*** 이상필, 한국과학기술정보연구원 책임연구원, 02-3299-6066, splee@kisti.re.kr

따라서, 과학기술정보를 효율적으로 활용하는 방법, 과학기술정보를 통해서 임무에 적합한 지식을 찾아내는 일이 점점 중요해지고 있으며, 과학 기술자들의 주요 관심 정보는 선진 연구기관의 연구동향과 미래의 과학기술 예측정보 등에 큰 관심을 보이고 있다. 그러나 이러한 정보는 실제로 알기가 쉽지 않으나, 지식기술 기법을 활용한다면, 정보의 흐름과 예측을 통해서 그들의 미래 연구 방향을 분석해 낼 수 있다. 이를 해결하기 위하여 과학기술 선진국을 중심으로 semantic, knowledge mining, informatics, informetrics, knowledge ontology, technology trend analysis, alert system 등의 다양한 분야의 정보분석 시스템 연구가 활발히 진행되고 있다.

이 중에서도 과학기술분야에서 계량정보학은 2000년대 이후에 급속히 연구되고 있다. 계량정보학은 과학기술문헌을 대상으로 어떤 특성을 도출하거나 혹은 특정한 연구문제를 해결하기 위한 수단으로서 과학문헌을 정량적으로 표현하는 접근을 지칭하는 용어로 계량서지학(bibliometrics), 계량과학학(scientometrics), 계량정보학(informetrics)이 혼재되어 사용되고 있다.

계량정보학이라는 용어는 계량서지학과 계량과학학의 영역을 포함하여 '정보'에 대한 정량적 접근을 통칭하는 학문분야 혹은 접근방식으로 사용하자는 독일의 학자에 의해 1979년 제안된 것으로 알려져 있다.

계량정보학의 연구범위에 대해서는 사용된 기법에 따라 빈도분석, 군집분석, 다차원 축적법 등으로 구분하거나 연구목적에 따라 성과 측정 연구, 구조와 맵핑 연구, 기술통계 등으로 구분하거나, 분석에 사용된 자료의 유형이 무엇인가에 따라 인용, 저자, 제목, 색인, 협력의 출처 등으로 구분하는 경우도 있다.

물론 과학활동에 대한 계량정보학적 접근에는 많은 한계점이 내포되어 있다. 계량정보학이 적용 가능한 과학활동의 범위가 출판된 논문이나 특허 등으로 제한되기 때문에 학자간의 비공식적인 지식의 흐름과 확산, 이를 통한 기여가 무시되고 있다는 점 등을 들 수 있다.

2. 기존 과학기술 정보분석 시스템의 한계

한편, 학문적인 방법으로 계량과학학, 계량정보학 등이 발전하면서 이를 실제 문제에 적용하기 위한 다양한 정보분석시스템 기술이 개발되었다. 대표적으로 미국 조지아텍대학의 VantagePoint, 오스트리아 ARC(Austrian Research Center)의 BibTechMon 등의 문헌정보 분석시스템을 들 수 있다. 이 외에도 문헌정보 중에서 특별히 특허정보를 특화시켜 분석하는 MicroPatent 등의 다양한 툴이 개발되고 있다.

그런데 1990년대 중반 이후부터 이러한 다양한 연구방법이 개발되었지만 실제로 과학기술정보를 통해서 연구개발의 문제를 해결하고 지식을 찾는 데는 한계를 지니고 있었다. 많은 연구와 관련 시스템이 개발되었지만 실제로 적용하여 활용하는 데는 많은 문제점이 있었던 것이다. 지금까지의 과학기술 정보분석 시스템 연구에서 가장 큰 문제점은 정보를 통해서 지식에 활용하는 과정에서 발생하였다. 즉 지금까지의 대부분의 연구는 계량서지학(bibliometrics)으로부터 출발하면서 주어진 정보자원을 어떻게 계량화할 것인가의 문제에 초점을 두고 있다. 그런데 실제로 필요한 지식은 정보의 계량화가 아닌 지식의 구조와 지식의 흐름을 밝히는 것이다. 즉 기존의 정보로부터 정보맵을 작성하는 구조적인 방법을 탈피하여 활용 가능한 지식맵을 밝혀내는 것이다.

또한 국내에서는 아직 정보자원의 구조적인 접근을 통한 계량화도 시스템적으로 개발하지 못하고 있다.

II. 과학기술정보분석시스템 동향

1. 과학기술 정보분석 시스템의 필요성

인터넷 환경의 급속한 발달과 최근 과학기술정보의 폭발적 증가로 이들 정보들 중 가치있는 정보를 받아 들이고 분석하는데 상당한 시간과 노력이 필요하게 되었다. 이러한 필요에 의해서 적절히 정보를 수집하고 분석하기 위한 다양한 기법들이 연구되고 있으며, 나아가서 이들 정보를 시스템적으로 분석할 수 있는 정보분석 시스템이 개발되고 있다. 이러한 과학기술 정보분석 시스템들 중 실제 활용되고 있는 미국 및 유럽에서 개발된 주요 정보분석 시스템의 개발 동향을 소개하고 나아가서 향후 과학기술 정보분석 시스템의 발전 방향을 제시한다.

2. 과학기술 정보분석 시스템 동향

1) Vantagepoint

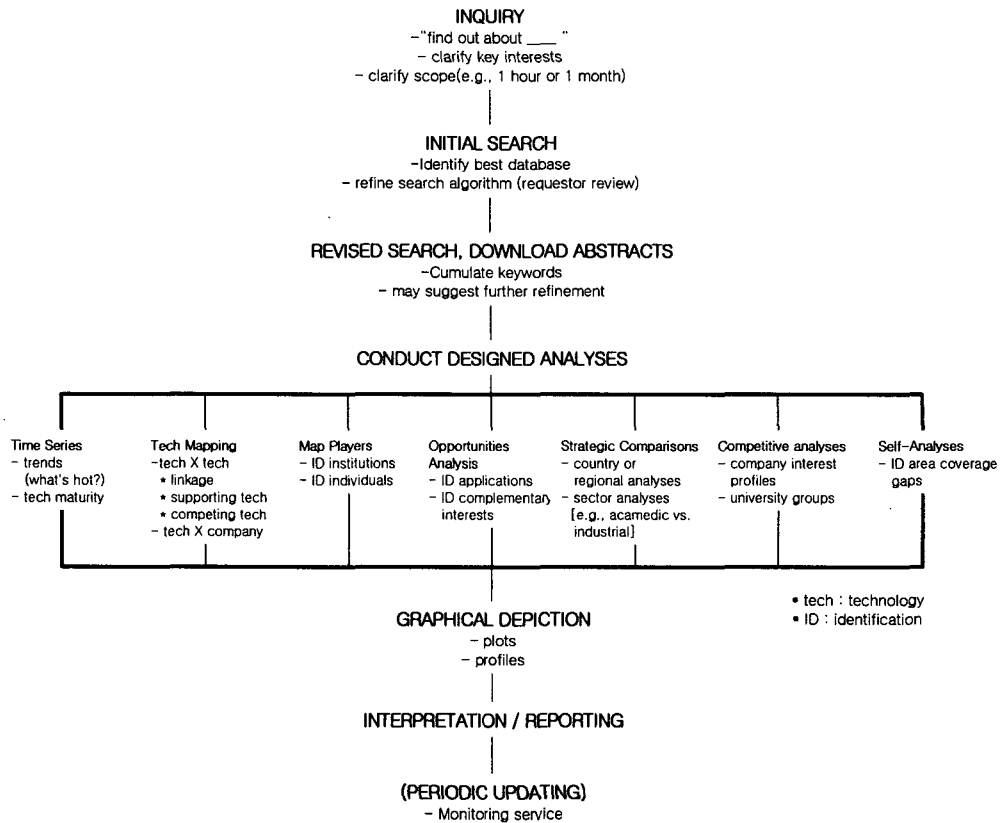
Vantagepoint는 미국 Technology Policy and Assessment Center of the Georgia Institute of Technology에서 개발한 TOAC를 Search Technology사가 상용화 한 것으로서, 과학관련 문헌들이나 특허 자료들 간의 구조적 상관관계를 정량적으로 분석하여 기술개발 동향과 활성도를 도출해내고 이를 도식화 하는 기능을 가지고 있다. Vantagepoint는 특정 주제에 대한 문헌DB에서 정보를 검색한 후, 용어 처리 과정을 거쳐 키워드를 추출하고, 추출한 키워드와 문헌들의 분류를 위해 주성분분석(Principal Components Analysis, PCA)과 같은 유도적 통계분석을 실행한다. 분석결과는 다차원척도법과 같은 다양한 종류의 연계, 클러스터링 방법 등을 이용하여 그래프로 표시된다.

분석 방법으로는 소스 데이터를 Import할 때 구분해 놓은 각 필드별로 리스트를 작성하고 리스트간의 상관관계를 분석하여 그래프화 하거나 매트릭스를 비교하는 방법, Co-occurrence, Auto-correlation, Cross-correlation, Factor map, PCA, PCD 등을 포함하고 있다.

Vantagepoint의 주요 사용 용도를 나열하면 다음과 같다.

- Scanning - identification of new technologies, developments in existing technologies, and new uses of technologies.
- Profiling - discovery of the key people and organizations.
- Mapping and Decomposition - identification of key dependency relationships among technologies (other technologies, scientific phenomena, manufacturing capabilities, etc...).
- Trending - establishing how a technology has emerged, its applications, and what factors (technical and non-technical) appear to govern its development.
- Forecasting - projecting how a technology could evolve, how it might diffuse into application, and the potential impacts of these events.

<그림 1> Vantagepoint의 상세 프로세스



2) BibTechMon

BibTechMon은 Co-word 분석에 의해 대량의 문헌 정보를 구조화 하고, 시각화 하는 지식맵 (knowledge map) 시스템으로서 Austrian Research Center에서 개발하였다. BibTechMon의 정보 분석 프로세스를 <표 1>에 나타내었다. BibTechMon은 CORDIS 데이터를 사용하여 EU연합에 가입하지 않은 동유럽 국가의 주력 연구분야 및 연구자를 찾아냄으로써 향후 연구 협력이 가능한 분야를 탐색하는데 적용된다. 특히 '정보화 사회 기술', '생활의 질과 주거자원'의 분야에서 연구협력이 가능한 분야를 찾아내는 것을 주요 목적으로 한다. Co-occurrence 분석에 의해 입력 받은 텍스트 정보를 자동 분류하고 키워드를 추출해서 knowledge map을 자동 생성하는 기능을 가지고 있다. 문서간의 상관성은 Spring 모델을 기반으로 하는 Iteration 알고리즘에 의해 구현되는 버블과 연결선으로 표현된다.

특히 특허 정보를 분석하는데 유용하지만, 기타 텍스트 정보를 내부 DB로 받아들이는 기능이 있기 때문에 문헌정보의 분석도 가능하다. 최종적인 분석 결과는 문헌, 주요 용어, 프로젝트 파트너, 특허출원인, 특허분류, 회사, 제품 등의 네트워크 그래프를 생성하는 것이다.

<표 1> BibTechMon의 정보 분석 프로세스

1. Database search
2. Build-up of a database containing relevant information
3. Automatic keyword generation
4. Generation of knowledge maps based on Co-word analysis
5. Cluster analysis

3) 기타 정보분석 시스템

국내에서 과학기술정보분석시스템은 주로 특허정보분석시스템을 중심으로 개발되었으며, 특허청의 PIAS(www.kipo.go.kr/www.patentmap.or.kr), Winslab의 INAS (www.winslab.com), WIPS의 PMS: Patent Management System (pms.wips.co.kr) 등을 들 수 있다.

또한, 네트워크 분석시스템으로는 UCINET, Pajek, NetMiner II, STRUCTURE, MultiNet 및 StOCNET 등을 대표적으로 들 수 있다. 이중 NetMiner II는 국내 벤처기업이 개발한 상용 네트워크 분석시스템이다.

3. 새로운 정보분석 시스템의 제안

혁신주도형 경제와 과학기술정책의 패러다임변화에 따라 우리나라는 과거의 추격적 전략(catch-up strategy)을 포기할 수는 없지만 동시에 선택적 집중(selection & intergration)으로 원천기술을 확보해야 하는 단계에 이르고 있다. 단순한 과학기술의 추적보다는 국가적 아젠다에 맞춘 전략적 지식의 창출과 혁신시스템(innovation system)의 구축이 필요하다¹⁾. 즉 제도적 지원을 통한 시스템의 기반위의 독창적 아이디어와 다원적 기술개발시스템이 필요하다.

이용자 측면에서 효율적 정보 활용을, 환경 변화 측면에서는 차세대 정보 분석 체계가 요구되고 있다. 이용자 측면에서는 효율적인 정보 활용을 통해서 연구 활동의 시간과 비용을 줄이고, 기존 지식을 강화하거나 새로운 지식을 예측할 수 있는 도구가 필요하며, 환경 변화 측면에서는 국내에서는 아직 이러한 분야에 대한 체계적인 연구가 진행되지 않고 있지만 그 어느 때보다도 정보 분석 체계의 구축에 대한 필요성이 시급한 실정인 것이다.

고급 분석 정보의 인프라 지원체제 구축을 통해 국가 R&BD 활동의 생산성을 제고하고, 연구 개발 활동 단계에서 많은 시간과 비용이 수반되는 기술 동향 분석 및 R&D 성과의 사업화를 위한 산업 시장 분석 정보를 공동 활용하는 지원 체제를 조속히 확립해야 할 것이다. 특히, 국가 차세대 성장동력산업의 핵심 기술 개발을 효과적으로 지원하기 위한 전문적인 정보 분석 및 제공이 효율적으로 이루어져야 할 것이다.

이러한 문제를 해결하기 위한 방법으로 계량정보학의 많은 연구가 진행되어왔다. 그러나 이러한 연구개발에도 불구하고 실제적인 지식맵을 작성하기 위한 시스템은 개발되지 못하고 있다. 따라서 기존 시스템의 문제점과 이를 해결하기 위해서는 새로운 접근 방법이 필요하다 하겠다.

먼저, 하이브리드접근에 의한 구조적 접근방법의 해결이 필요하다. 앞서 연구배경에서도 언급했듯이 실제적인 수요자의 요구를 반영하기 위한 정보분석시스템은 제대로 개발되지 못하고 있다. 지금까지는 정보를 데이터로 활용하여 처리하고 매핑하는 방법에 치우쳐있었던 것이다. 따라서 정보자원수집->정보의 구조화->정보맵작성->새로운 지식활용의 4단계에서 정보맵 작성을 통해서 실제로 정보를 전략적으로 활용하는데 많은 문제점들이 노출되었다. 즉 활용목적에 포커스를 맞춘 것이 아니고 정보를 처리하여 맵을 작성하는 부분 중심으로 연구되었고 대부분의 시스템도 이러한 분야에 국한되었다. 그렇다보니 실제로 정보를 전략적으로 활용하는 데는 큰 효용가치를

1) 오세정, 이달곤, "과학기술혁신체제의 재구축과 전개방향", 국가과학기술혁신체제 개편에 관한 공청회, 2004.3

느끼지 못하게 되었다. 최근 포터(*)는 Tech Mining이라는 보고서를 통해서 이 부분에 대한 문제를 본격적으로 다루기 시작하고 있다. 즉 정보 활용맵을 통해서 요구하는 문제를 해결하기 위한 로직을 개발하고 이에 따른 지식의 구조화 정보자원의 활용이라는 접근방식이 필요하다.

예를 들면 급부상하는 연구영역의 도출, 공백기술의 발견, 유망연구영역의 도출 등의 문제에 대해서는 개별적으로 정보맵을 통해서 해결하는 과정이 필요하다.

두 번째는 국내에서는 과학기술 정보분석시스템이 개발되지 못하고 있다는 것이다. 이러한 시스템의 국산화도 시급한 문제이다. 현재 기존의 시스템의 가격이 \$7,000 - \$18,000에 이를 정도로 고가이므로 이를 국내 연구자나 연구관리자가 사용하기는 힘든 실정이다.

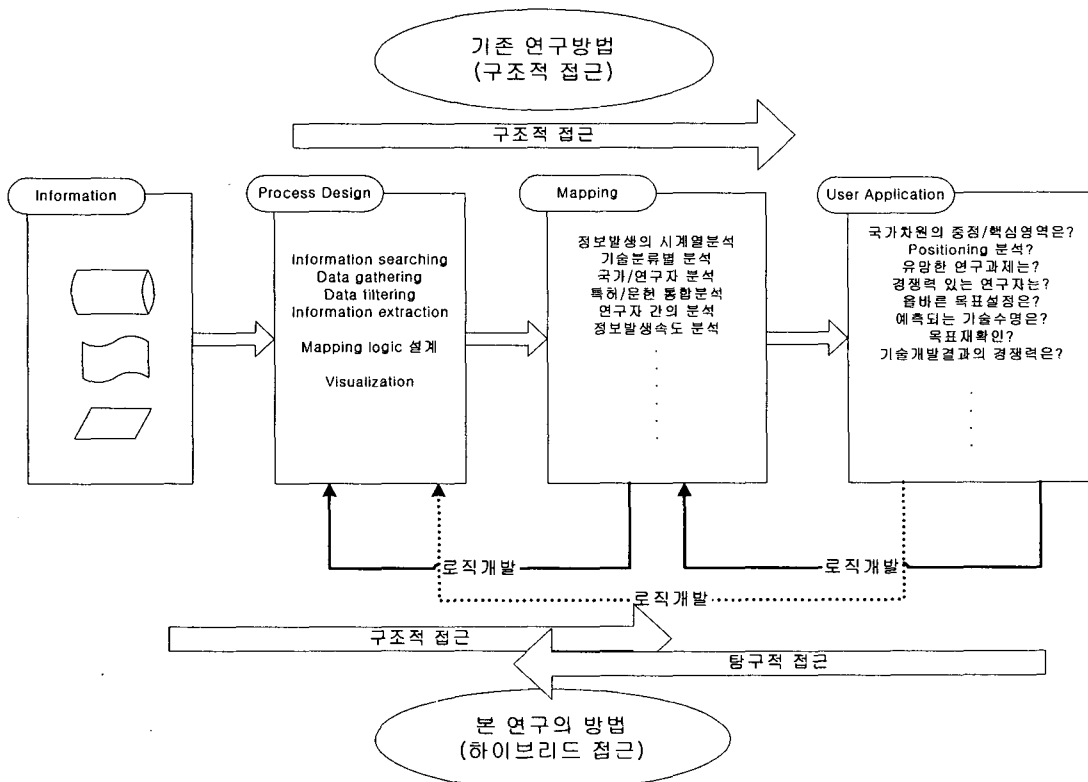
셋째는 기존의 개발된 시스템이 연구자 등의 지식네트워크 기능이 연계되어 있지 않다는 점이다. 연구활동은 연구결과로 나타나는 논문과 특허도 중요하지만 이를 수행하는 단계의 연구자 네트워크, 지식의 흐름, 지식간의 클러스터 등의 지식네트워크에 대한 정보맵도 밝혀져야 한다.

넷째는 alerting 기능이 해결되어야 한다. 수많은 쏟아지는 정보 중에서 연구자의 포지셔닝을 알고 있어야 하고, 경쟁적인 연구에 대한 체계적인 분석이 지원되어야 한다.

마지막으로 개인이 수집한 다양한 정보를 개인이 모델링하여 시뮬레이션할 수 있는 기반이 마련되어야 한다. 주어진 조건의 구조적인 문제, 즉 주어진 정보자원, 주어진 정보맵 작성방법으로는 수요자의 다양한 지식문제를 해결하기 힘들다. 따라서 시스템 수요자가 시스템을 개인화하여 필요한 정보를 구축하고 필요한 로직을 구성하여 정보분석할 수 있는 시스템개발이 필요하다.

따라서 새로운 과학기술정보분석시스템은 수요자 중심의 지식활용시스템에 초점을 맞추어야 할 것이다. 기존의 시스템이 정보분석에 치우쳐있지만 새로운 시스템은 정보분석, 평가 및 예측이 가능한 지식 추구형 시스템이 되어야 하겠다. 이를 위해서는 기존의 구조적 접근방법에 수요자의 지식활용에 적합한 탐구적 접근방법을 동시에 적용하는 하이브리드형의 시스템적 접근방법이 필요하다 하겠다.

<그림 2> 하이브리드형의 시스템적 접근



III. 과학기술 정보분석 시스템의 발전방향

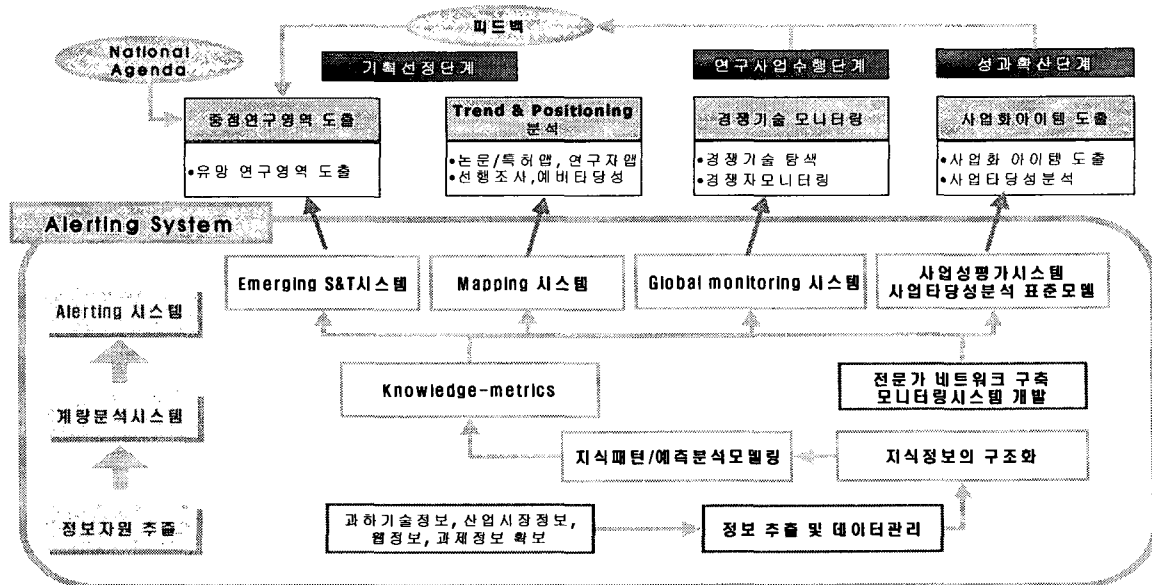
우리나라의 국가 과학기술 정책 방향이 선진국 모방형 기술 개발에서 원천기술 개발로 전환하는 과정에서 관련 연구개발 동향, 전략 기술, 산업에 대한 선행조사 및 분석의 필요성이 증가하고 있다. 또한, 국가 연구개발에 대한 전략성 확보, 사전기획·조정 및 조사·분석·평가 기능의 전문성·투명성 제고를 통해 국가연구개발사업의 기획 및 종합조정 기능의 내실화를 지원할 수 있는 과학기술 선도형 정보분석 시스템을 요구하고 있다. 이러한 국가적 니즈 및 수요지향적 연구개발을 지원하고 나아가서 국가연구개발사업의 효율적이고 생산적인 연구기획 강화를 위해서는 상시적인 R&D 전략정보분석 시스템의 개발이 시급하다 할 수 있다.

국가적으로는 NTIS(국가과학기술종합정보시스템) 구축 등 국가연구개발시스템의 혁신으로 이에 부응할 수 있는 지식 정보분석 시스템의 개발이 필요하며, 정보의 기하급수적 증가로 R&D 활동에서 글로벌 지식정보를 전략적으로 분석하여 지식정보의 고급화 및 활용가치를 극대화할 필요성도 매우 높아지고 있다. 지식정보의 고급화와 활용가치 극대화를 통해 R&D의 생산성 증대, 성공율 제고 및 연구성과 품질 제고를 유도함으로써 R&D 투입자원의 효율성을 높일 수 있다는 것이다.

최근의 국제적인 연구동향을 살펴 보면 글로벌 동향의 분석(global monitoring), 지식계량분석(knowledge-metrics), 기술의 위치와 수준분석(technology mapping), 폭발적으로 증가하는 정보에서 가치 있는 새로운 부가가치 지식을 창출하는 KDD(knowledge discovery in database)기법 등이 정보분석의 주요 연구과제들로 대두되고 있다. 이러한 정보분석과 관련된 기술의 발전에 적극적으로 대처하기 위해서는 패러다임 변화를 적극적으로 수용하여 문제 해결뿐만이 아니라 문제를 적극적으로 제기할 수 있는 능동형 정보분석시스템의 개발로 나아가야 할 것이다.

이러한 시대에 부응하는 과학기술 정보분석 방향을 지원하기 위해서는 이를 시스템적으로 해결할 수 있는 차세대 과학기술 정보분석 시스템이 구축되어야 할 것이다. 과학기술 지식자원이 2%밖에 생산되지 않는 우리나라로서는 전세계의 지식을 스크리닝, 모니터링하고 분석하여 연구방향과 연구수행에 적절히 반영할 수 있는 시스템의 개발이 필요한 것이다.

<그림 3> 국가 과학기술 정보분석 시스템(안) 개요도



[참 고 문 헌]

- [1] 문영호 외(2000), 《온라인 DB 검색을 통한 기술분석시스템 구축》, 산업자원부.
- [2] 김경호·문영호 외(2001), 《문헌정보분석을 통한 기술예측 시스템의 개발》, 공공기술연구회.
- [3] 유사라(1999), 《정보학연구와 분석방법론》, 나남.
- [4] 문영호 외(2005), 《차세대 과학기술 정보분석/예측/평가 시스템 개발》, 공공기술연구회.
- [5] 설성수(2002), “기술분석의 고도화”, 기술혁신학회지, Vol. 5, No. 3, pp. 260-276.
- [6] Hood, W. W.(1998), *An informetric study of the distribution of bibliographic records in online databases: a case study using the literature of Fuzzy Set Theory (1965-1993)*, Sydney: Ph.D. dissertation, The University of New South Wales.
- [7] Narin, F.(1994), “Patent Bibliometrics”, *Scientometrics*, Vol. 30, No. 1, pp. 147-155.
- [8] Mark Huisman, Marijtje A.J. van Duijn(2005), “Software for social network analysis”, New York: Cambridge University Press.
- [9] <http://www.thevantagepoint.com/index.html>(Search Technology사의 Vantagepoint 홈페이지)