

기술문헌DB로 부터 기술동향분석을 위한 기술정보 분석시스템 개발에 관한 연구

Development of Technology Information Analysis System for Analyzing Technology Trend from Technology Data

이호신
 한국과학기술정보연구원 정보분석부 선임연구원
 이일형
 한국과학기술정보연구원 정보분석부 선임연구원
 김경호
 한국과학기술정보연구원 정보분석부 책임연구원

Ho-Shin Lee
 Senior Researcher, Dept. of Information Analysis, KISTI
 Il-Hyung Lee
 Senior Researcher, Dept. of Information Analysis, KISTI
 Kyung-Ho Kim
 Director, Dept. of Information Analysis, KISTI

중심어 : 데이터베이스, 정보분석, 기술동향, 분석시스템

요약

최근 지식기반사회에 접어들면서 정보의 양은 폭발적으로 증가하고 있으며, 인터넷과 디지털 콘텐츠의 확산으로 인해 정보의 접근성 또한 매우 향상되었다. 이로 인해 세계적으로 방대한 정보의 체계적인 관리와 효율적인 분석이 큰 이슈가 되고 있다. 또한, 정보의 패러다임이 정보의 접근시대에서 정보의 분석시대로 전환되고 있으며, 이에 대비하여 기술정보를 통한 기술트렌드의 분석과 이를 시스템적으로 구현하는 방법의 정립이 시급한 실정이다. 따라서, 본 연구에서는 기술정보 DB의 검색결과를 온라인상에서 시스템적으로 신속하게 분석함으로써 연구개발 및 기술개발의 방향설정에 기여하며 정보수집 및 분석에 소요되는 노력과 시간을 획기적으로 단축시킬 수 있는 정보분석시스템을 개발하였다.

Abstract

As the knowledge-based society progresses the amount of information increases tremendously and also the accessibility to information is much improved through the diffusion of internet and digitalized contents. At this juncture the systematic management and efficient analysis of immense information surfaces as a global issue. And the paradigm of information turns from the age of information access to the age of information analysis. So it is urgently required to analyze technology trend from technology information and to establish the way to realize a systematic method for information analysis. Accordingly in this study an information analysis system is developed that contributes to the decision of proper direction for R&D and technology development by rapidly and systematically analyzing search results from the technology databases online and can reduce to a great degree the efforts and time required to collect and analyze information.

1. 서론

2002년 들어 세계경제는 선진국들의 동반 침체와 함께 둔화세로 전환되었으며, 특히 미국경제는 IT산업 침체의 여파로 성장률이 빠른 속도로 하락하였고, 일본 또한 극도의 내수침체속에 성장률이 마이너스로 하락하는 어려움을 겪고 있다. 그러나, 이러한 세계경제의 침체속에서도 인터넷

을 기반으로 한 정보규모는 지속적인 성장을 거듭하였으며, 지식기반사회에서 다양한 형태의 정보가 사방에서 제공되고 있다[1]. 미래에는 이러한 무수한 정보중에서 자신이 원하는 정보를 찾아내기가 더욱 어려워질 것이다. 이른바 정보의 홍수속에서 정보를 효율적으로 수집하고, 이를 효과적으로 활용하는 사람(조직)과 그렇지 않은 사람 사이에는 매우 큰 격차가 생기게 된다. 개인과 조직, 그리고 사회의 존

재와 활동이 정보에 크게 의존하게 되고, 필요한 정보를 누가 얼마나 신속하고 정확하게 확보하여 분석하고 미래를 예측하느냐에 따라 경쟁의 우열이 결정될 것이다. 특히 정보검색 및 분석의 문제를 경쟁우위를 확보할 수 있는 『열쇠』로 인식하고 이를 가장 적극적으로 다루는 개인, 기업 및 정부야말로 세계경제의 침체속에서 세계 일류로서의 자리를 쟁취할 수 있을 것이다. 현재의 기업 환경은 격렬한 기술개발 경쟁, 신제품의 라이프사이클 단축, 시장개척 및 유지의 어려움, 자유무역 체제로의 전환, 지구 환경의 보존을 위한 규제강화 등에 의해 매우 복잡해지고 있으며, 의사결정을 어렵게 하는 불확실성의 정보는 시대의 변화와 함께 더욱 커지고 있다. 이러한 환경의 변화속에서 기업이 살아남기 위해서는 경쟁자와 차별화 할수 있는 창조적인 능력과 환경 변화를 신속하게 예측하고 대응방안을 모색할 수 있는 기능을 강화하지 않으면 안되며, 이를 위해서 정보의 분석 및 활용은 가장 기본적인 요소이다. 정보전달이 주목적인 콘텐츠를 계량적인 방법으로 분석하여 기술분석의 관점에서 활용하고자 하는 것은 정보와 기술개발은 하나의 대응적인 사상을 발생시키기 때문에 가능한 것이다. 직접적인 정보분석을 통해서 기술의 흐름을 파악하는데 응용할 수 있다는 이러한 관점의 접근은 정보(또는 DB)의 각각의 데이터는 연구성과나 혁신의 성과, 아이디어의 산물로서 표현된 것이고, 이러한 정보의 흐름을 분석하는 것은 연구의 흐름을 분석하는 것이 된다. 또 DB를 통하여 필드간의 관계분석, citation 분석 등을 통하여 연구의 동향과 기술분석에 적용할 수 있다는 점이다.

본 논문의 구성은 다음과 같다. 2장에서는 기존 연구내용과의 차별성에 대해 소개하고, 3장에서는 정보분석시스템 모듈 설계의 기본이 되는 DB 필드 구조와 각 항목에 대한 분석인자 추출에 관하여 논하였고, 이와 함께 분석기법의 소개 및 분석인자의 적용방안에 관하여 서술하고, 4장에서는 이의 시스템화를 위한 분석모듈, 검색 및 DB생성 모듈, 콘텐츠 관리모듈 설계에 관하여 기술하고 결론을 맺는다.

II. 기존 연구

과거에는 정보의 수집이 제일 중요한 정보활동이 되었지만 이제는 정보의 수집과 더불어 정보의 효과적인 분석이 주요 관심이 되고 있다. 이미 정보의 효율적인 분석방법, 새로운 기술변화를 예측하는 정보예측 방법에 대한 연구가

일본, 미국 등의 선진국들에서는 중요한 과제로 대두되고 있다[2]. 기술정보의 분석을 위한 방안으로는 기술정보 DB를 통해 검색한 정보를 DB 필드의 정량적 분석 등의 기법을 이용하여 시스템적으로 분석하여, 기하급수적으로 늘어나고 있는 정보흐름과 정보가 내재하고 있는 기술의 트렌드를 파악하여 DB를 통해서 새로운 정보와 연구, 기술개발의 트렌드를 파악할 수 있다. 미국, 일본 등 외국에서는 1990년대 들어서면서부터 온라인으로 서비스되고 있는 기술DB를 검색하고, 이러한 2차 정보를 이용하여 정보분석을 수행하는 연구와 실용화가 부분적으로 이루어지고 있기는 하나[3][4] 국내에서는 기술정보의 분석기법에 대한 현황조사나 기본적인 연구가 체계적으로 이루어진 사례는 거의 없다. 다만, DB를 통해 검색된 결과를 분석하는 방법 중에서 특허정보의 경우 제한적인 형태의 정량적 분석의 연구가 진행되고 있으며 이를 이용한 제품이 출시되고 있다[5]. 그러나 특허 이외의 기술문헌 정보DB를 분석하는 시스템은 전혀 구축되어 있지 않은 실정이다. 따라서, 본 논문에서는 국내 최초로 기술문헌 정보DB를 활용하여 온라인상에서 기술 수준 및 추이를 분석할 수 있으며, 정보분석에 소요되는 시간 및 노력을 획기적으로 단축시킬 수 있는 시스템의 설계를 제안하였다.

III. 정보분석용 DB 및 분석인자 설계

1. DB 선정

본 연구에서는 한국과학기술정보연구원(KISTI)이 보유·운영중인 기술문헌관련 DB를 대상으로 하였다. 표 1에 기술한 바와 같이 기술문헌관련 DB로는 COMPENDEX, INSPEC, MCAT, DIMD, SOCIETY를 시스템 구축에 활용하였다. 본고에서는 지면상의 제한으로 대표적인 과학기술문헌 DB인 COMPENDEX를 중심으로 기술하였다.

표 1. 정보분석에 활용한 DB (2002. 7월 기준)

DB명	수록정보	수록기간	수록규모	비고
COMPENDEX	공학일반	1970 -	4,684,678	해외도입
INSPEC	전기전자	1969 -	7,321,047	해외도입
MCAT	연구보고서		55,513	KISTI구축
DIMD	학위논문		651,850	KISTI구축
SOCIETY	학회논문		150,578	KISTI구축

2. 항목 분석

COMPENDEX는 공학일반을 다룬 DB로 미국의 Engineering Information Inc에서 제작하였으며, 현재 세계 각국에서 만들어지고 있는 4,000여종의 온라인 데이터베이스 중에서 가장 대표적으로 DB중 하나이다. 총 14개의 필드 항목으로 구성되며 각각의 항목은 일정한 규칙에 따라 작성된다. COMPENDEX 콘텐츠의 건본은 그림 1에 나타내었으며, 구체적인 필드 항목에 대한 분석 내용은 다음과 같다.

DN : COM1999494864423
IS : 199949
DT : Journal Article
LA : English
TI : Hierarchical interconnection network architecture
AU : Orenck, Bulent
CS : Istanbul Technical Univ Istanbul Turk
SO : Turkish Journal of Electrical Engineering and Computer Sciences v 6 n 2, p131-166, 1998, 56 Refs.; TR: T;
AS : Pub: Sci Tech Res Counc Turkey Ankara Turkey
SN : 1300-0632
CC : 717.1; 722.4; 716; 716.1; 716.3; 716.4;
KW : Multiprocessing systems; Client server computer systems; Real time systems
ID : Optical interconnects
AB : A hierarchical network suitable for interconnection of real-time client processes in a distributed multiprocessor environment is presented in this paper. A multi-layer Communication Unit (CU) prototype is developed for this purpose. This unit offers the client processes communication services for real-time operation.

그림 1. COMPENDEX 콘텐츠 건본

① DN : 문헌번호(Document Number)

데이터베이스명과 일련번호로 구성되어 있으며, 일반 이용자에게는 의미가 없으며 분석인자로 사용되지 않는다.

COM1999494864423 은 COM : COMPENDEX, 199949 : IS Number, 4864423 번째 레코드라는 의미이다.

② IS : COMPENDEX의 권·호(Volume, Issue)

COMPENDEX의 권·호를 표시하며, YYYYNN의 형태로 되어 있다.

YYYY : 연도, NN : Serial Number

③ DT : 문헌의 형태(Document Type)

원문의 형태를 나타내며, 1984년 이전 자료에 대해서는

분류가 되지 않고 있다. DT 항목은 기술개발 수준분석에 활용하게 되며, 저널 논문, 컨퍼런스 논문, 리뷰 보고서 등의 상대 발생 비율 및 증가율을 분석인자로 활용하게 된다.

J : Journal Article

C : Conference Article, Conference Proceedings

R : Report Review

M : Monograph Chapters

P : Preprint

④ LA : 언어(Language)

원문의 언어를 나타내며, 1984년 이전 자료에 대해서는 분류가 되지 않고 있다.

⑤ TI : 제목(Title)

기사의 원제목이 실질적으로 이용된다.

⑥ AU : 저자(Author)

저자명이 수록되는 형태는 원문에 주어진 정보에 따라 달라지나 GUNDERSON, J.A와 같이 성, 콤마, 첫 번째 이름의 대문자, 도트, 두 번째 이름의 대문자, 도트로 구성된다.

AU 항목은 기술전문가 분석에 활용되며 방대한 양의 정보를 분석하여 최다 저자 및 연도별 저자수 등의 분석이 가능하다.

⑦ CS : 저자의 소속기관(Corporate Source)

표 2. 국가코드명

국가명	코드	국가명	코드	국가명	코드
미국	USA	스위스	Switz	호주	Aust
일본	Jpn	싱가포르	Singapore	이탈리아	Italy
중국	China	독일	Ger	러시아	USSR Russia
캐나다	Can	아일랜드	Irel	이스라엘	Isr
프랑스	Fr	대만	Taiwan	한국	Korea
영국	UK	오스트리아	Austria	아프리카	Afr
스웨덴	Sweden	이집트	Egypt	스페인	Spain

원문에 수록된 소속기관명이 채택되거나 다수의 저자가 여러 기관에 소속되었을 때는 첫번째 저자의 소속기관명이 채택된다. 기관명 영역에 수록되는 데이터는 조직의 명칭과 도시, 주 또는 지방, 국기를 포함하는 조직의 위치가 수록된다.

CS 항목은 관련기술의 주요기관 분석 인자로 활용되며 국가별 주요기관 분포 및 현황 분석도 가능하다.

⑥ SO : 서지사항(Source)

잡지명, 자료의 권(V ; Volume), 호(N ; Number), 면수, 발행년도 ISBN(국제표준단행본번호), TR(처리 코드)로 구성된다. SO 항목은 주요저널 분석인자, 연도 인자, 기술수준 분석인자로 크게 세가지 분석인자를 포함하고 있다. 따라서, 효과적으로 각각의 인자를 구분하여 서로의 상관관계를 통계적으로 분석할 수 있는 설계가 요구된다.

처리 코드(TR)는 A, B, C, G, N, P, T, X로 8개 분류로 부여된다. 특히 A와 P는 응용기술관련 콘텐츠, G는 기술리뷰관련 콘텐츠, T는 이론적 연구관련 콘텐츠, X는 실제 실험적 결과 관련 콘텐츠로 이들 5개의 처리코드는 기술의 발전단계를 분석하는 데 활용되게 된다.

표 3. 처리코드(TR : Treatment Code)

코드	내용
A	Application : 물리적 효과 방법 장치 등의 실제
B	Bibliography : 문헌의 도서목록
C	Commercial : 가격개발 등에 대한 예측
G	General : Review 나 State of the Art
N	News : 새로운 효과, 장치기술
P	Practical : 직접적인 실제사용
T	Theoretical : 이론적, 수학적 방법으로 처리
X	Experimental : 실험방법, 관찰 등의 결과에 관한 사항

⑨ AS : 부가 서지 사항

발행처, 발행국등

⑩ SN : 국제표준연속간행물번호(ISSN)

⑪ CC : 주제분류코드(Classification Codes)

표 4. 주제분류코드(Classification Codes)

분야	그룹	분류범위
토목, 환경, 지질, 생체공학	400~480	401~484
광산, 금속, 석유, 연료공학	500~540	501~549
기계, 자동제어, 핵, 우주, 항공공학	600~690	601~694
전기, 전자, 제어공학	700~750	701~753
화학, 농학, 식품공학	800~820	801~822
산업공학, 경영, 수학, 물리	900~940	901~944

CC 항목은 각각의 콘텐츠에 대한 기술분류를 나타내는 코드로 정보분석에 매우 중요한 인자중에 하나이다. 400계

열부터 900계열까지 크게 6개 기술계열로 분류되며 다시 이들 대분류는 약 900개의 세분류로 다시 나누어 진다.

기술분류 코드분석을 통하여 특정기술의 상관관계 분석이 가능하며, 각 분류코드별 추이를 통하여 핵심유망기술을 도출할 수 있다. 또한, 기술분류코드에 대한 뉴 엔트리 분석을 통하여 기술의 파급성 분석이 가능하다.

⑫ KW : 색인어(Keyword)

색인어는 문헌의 제목과 초록에서 수정없이 추출되며, 다음과 같은 의미를 갖는다.

- 색인시 통제용어를 사용할 수 없거나 통제용어보다 주제에 대해 세분화된 명확성을 부여
- 저자가 사용한 일반 언어로의 검색을 부여
- 새로 발생된 신규 어휘

KW 항목은 추가된 색인어 및 출현 빈도수 등을 분석할 수 있는 인자로 뉴엔트리 분석이 주로 사용된다.

⑬ ID : 주제어, 식별어(Identification)

SHE(Subject Headings for Engineering)에 수록된 12,000 용어의 결합이나 개별용어의 통제색인어를 수록한다.

⑭ AB : 초록(Abstract)

AB 항목은 기술문헌에 대한 초록으로 직접적인 분석인자로는 활용이 불가능하며 사용자 정의 기능을 통하여 성능인자 추출 및 분석 항목 추가시 참고 자료로 활용할 수 있다.

3. 분석 기법 적용

3.1. 기술 수준 분석

전통적인 통계해석기법으로서 일반적으로 문헌의 수량을 파악하여 분석하는 방법이다. 일반적으로 건수추이 맵이라고도 하며, 직교좌표상에 년, 월 등의 시간축이 횡축으로 존재하고, 반면 종축에는 건수 등이 표시된다. 자세한 내용은 다음과 같다.

(1) 연도별 논문건수 및 누적 논문건수

가장 개괄적이고 일반적인 내용을 담고 있는 분석인자로서 연구대상 기술이 현재 어느 정도의 연구단계에 있는지 짐작할 수 있다. 또한, 누적논문건수에 대해서는 그림 2에 나타낸 바와 같이 외삽법을 적용하여 신기술 발생, 기술혁신, 대체기술 등장 유무 및 시기 등을 추정할 수 있다.

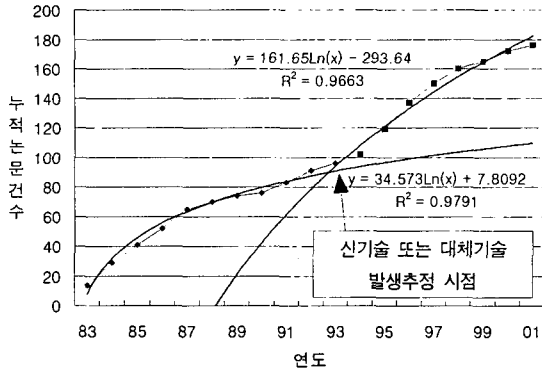


그림 2. 누적 논문건수 분석에

(2) 연도별 국가논문 건수

세계 각국의 기술수준이 어느 정도 연구단계에 와 있는지를 파악할 수 있다. 이 방법은 COMPENDEX DB 중 CS(Corporate Source : 저자의 소속기관) 필드의 마지막 부분에 표시되는 국가명을 기초로 하여 분석한다.

따라서 Journal Article 이외의 자료의 경우 CS 부분에 국가명이 없는 경우가 있으므로 사용자가 별도로 국가명을 추가하여 분석해야 한다. 또한 DB 구성내용의 CS 항목에서 설명한 바와 같이 출원빈도가 높은 국가만을 표시하고, 그 이외의 국가는 기타로 처리하여 분석하므로 이에 대한 오차가 발생할 수 있다.

(3) 연도별 저자 소속기관 건수

해당기술에 대한 국내외 연구단계를 파악할 수 있다. 즉, 국내외 연구소 중에서 어느 기관이 해당기술에 대해 활발한 연구를 진행하고 있는지를 파악할 수 있다.

(4) 연도별 저자 건수

해당기술에 대한 국내외 저명한 연구원들을 파악할 수 있으며, 연구과제의 보편화도, 성숙도 등을 판단할 수 있는 지표가 될 수 있다. 특히 (2), (3)의 결과와 병행하여 분석할 경우 국가와 저자 및 소속기관들에 대한 기술 격차 등 해당기술에 대한 상관 관계를 분석할 수 있다[6].

(5) 각 연도별 국가 비율

임의의 한 국가 문헌이 세계에서 차지하는 비율을 추적함으로써 기술격차 및 연구개발의 진척도 등을 파악할 수 있다. x축을 연도로, y축을 $\frac{\text{해당 국가의 문헌 수}}{\text{전체 문헌 수}}$ 로

설정하고 꺾은선 그래프를 작성한다.

(6) 연도별 문헌형태 비율

연구과제의 성숙도 등을 파악할 수 있는 인자로서, Document Type의 비율에 따라 해당 기술의 현단계 수준을 가늠할 수 있다. 즉 순수 논문의 성격을 가진 Conference 자료가 많은 경우 해당기술의 수준이 연구실 단계로 볼 수 있으며, 또한 리뷰 기사의 성격을 가지는 Journal Article 자료가 많은 경우 기술 수준이 보편화되어 있다는 것을 알 수 있다.

(7) 포트폴리오(Portfolio) 맵

이 분석인자는 특허분석에서 많이 쓰이는 분석인자로서 그림 3과 같이 나타낼 수 있다[7]. 포트폴리오 맵은 일정투자에 따른 최대의 수익을 얻기 위한 수익성 분석의 경제용어이나 이를 기술문헌정보 분석에 응용한 것이다. 횡축인 x축에 최근 수년간(3-5년)의 논문건수를 표시하고 종축인 y축에 논문건수의 신장율을

$$\left(\frac{\text{당해연도의 논문건수} - \text{전년도의 논문건수}}{\text{전년도의 논문건수}} \times 100 \right)$$

을 표시한다. 또, 누적논문건수를 원의 형태로 크기에 따라 표시한다. 이 같이 표시된 그림을 서로 연결하면 좌우축 변수에 따라 해당기술의 라이프사이클이 어떠한지 또는 현재 그 기술의 발전형태는 어떤 상태인지를 예측할 수 있다.

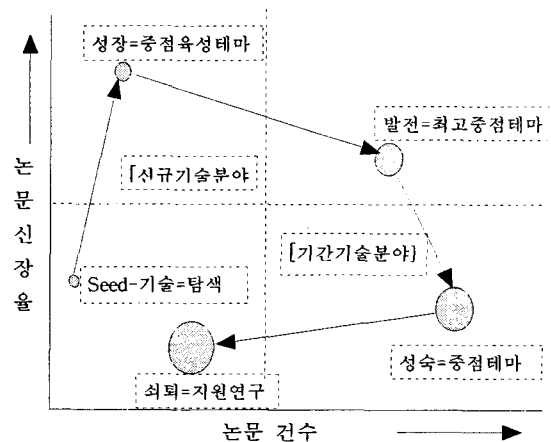


그림 3. 포트폴리오 맵

기술의 발생과 소멸의 관계를 중형축의 관계와 비교하여 설명하면 다음과 같다.

- 탐색기(신기술의 탄생)
 - 논문의 발생건수가 적고 신장율이 낮다.
- 성장기(신기술의 성장단계)
 - 논문의 발생건수가 증가하고 신장율이 상승한다.
- 발전기(기술의 발전단계)
 - 논문의 발생건수가 많고 신장율이 둔화된다.
- 성숙기(기술의 성숙단계)
 - 논문의 발생건수가 많고 신장율이 저하된다.
- 쇠퇴기(기술의 쇠퇴단계)
 - 논문의 발생건수가 감소하고 신장율이 저하된다.

(8) 각 연도의 논문건수 vs. 각 연도의 저자수

그림 4에서 처럼 Patent Mapping 기법(건수추이 맵)을 기술 문헌에 적용한 것으로 건수의 추이에 따라 기술의 발전기, 성숙기, 퇴조기, 부활기를 파악할 수 있다.

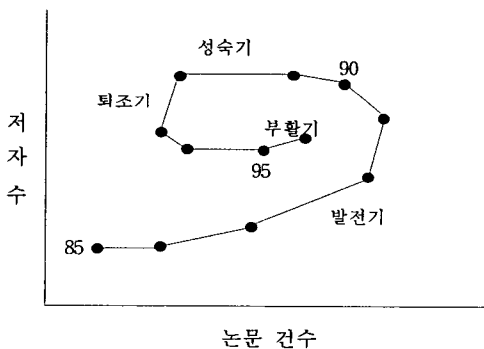


그림 4. 각 연도의 논문건수 vs. 각 연도의 저자수 맵

3.2. 기술 분야 분석

COMPENDEX DB의 경우 처리코드, 주제별 분류 코드, 주제어 등이 부여되어 있어 각 논문의 기술분야를 일일이 사용자 읽어서 구분할 필요없이 위와 같은 코드를 이용하여 분석할 수 있다. 이를 이용한 분석의 자세한 내용은 다음과 같다.

(1) 연도별 처리코드 분포

처리코드는 데이터베이스 구성내용에서 설명한 바와 같이 문헌의 성격을 나타내기 위해 8가지의 코드를 부여하였다. 따라서 문헌 성격의 종류에 따라 해당기술의 수준을 파

악할 수 있다.

(2) 연도별 분류코드 분포

COMPENDEX의 주제분류코드는 앞서 언급한 바와 같이 상당히 자세하게 대분류, 중분류, 소분류로 나누어져 있어 기술간의 상관관계 등을 파악할 수 있다. 이 분석인자를 통하여 기술의 분야별 이동상황과 기술의 적용분야를 파악할 수 있다.

(3) 연도별 추가된 분류코드

대상기술이 어떠한 새로운 분야에 파급효과를 보이며, 또한 적용되고 있는가를 판단할 수 있는 기준이 된다.

(4) 연도별 색인어 출현 빈도 수

주요 색인어를 분석함으로써 대상기술을 보다 세분류하거나 신규 색인어를 파악하여 기술의 세부 적용 분야를 알 수 있다.

(5) 연도별 주제어 출현 빈도 수

COMPENDEX DB에서 부여된 주제어를 분석하는 인자로써 (4)와 유사한 방법이다. (2)의 연도별 주제분류 코드 분포와 성격이 비슷하므로 병행하여 사용할 경우 보다 더 유용한 분석결과를 얻을 수 있다.

(6) 연도별 잡지 게재 건수

잡지의 게재 건수를 파악할 경우 대상기술의 전문 잡지 등을 파악할 수 있으며, 이를 통한 기술적용 분야의 변천 및 확산 정도를 파악할 수 있는 보조적 기능을 한다. 단, 이 분석인자는 DB 특성상 문헌 형태가 Journal article인 경우만을 한정하여 분석하므로 이에 대한 오차가 발생할 수 있다.

3.3. 정성분석

기술문헌 DB를 분석함에 있어서 반드시 분석되어야 할 인자임에도 불구하고 필드가 설정되어 있지 않은 경우 사용자 정의를 통해서 분석해야 하며 이를 정성분석이라 한다. 프로그램 상에 사용자 정의 필드를 설정하여 사용자가 임의로 필드를 작성, 분석한다. 다음에 몇 가지 예를 제시하였다.

(1) 기술분야 맵

여기서의 기술분야 맵은 분류코드 상의 내용이 아닌 특 정과제에 대한 세분류를 의미하며 사용자가 편리한 대로 세분류하는 것을 의미한다. 3차원 그래프인 경우 x축은 연 도, y축은 기술분야 사용자 정의(예 : 제조법, 장치, 반응 메 커니즘, 특성분석), z축은 각 분야의 해당연도의 건수를 도 시한다.

(2) 기술 중요도 맵

사용자가 자신의 연구분야에 비추어 참고문헌의 중요도 를 임의로 평가해서 3차원으로 도시한다. 요지맵과 병행하 여 참고문헌을 체계적으로 정리하는데 도움을 줄 수 있다. x축에 연도, y축에 기술 중요도 (A, B, C, D, E), z축에 각 중요도별 해당연도의 건수를 표시한다.

(3) 요지 맵

각각의 문헌에 대하여 요지를 간략하게 사용자가 입력하 여 참고문헌의 체계적 관리를 가능하게 한다.

IV. 시스템 개발

1. 시스템 설계

웹기반하에서 대용량의 정보를 서버로 업로딩하여 데이 터를 분석하고, 기술추이를 그래프화하여 클라이언트에게 그 결과를 회송하도록 구성하려면 최적의 네트워크 환경, 펜티엄III, RAM 256M이상의 시스템 환경 등이 요구된다. 먼저 본 시스템의 사전 테스트 및 성능 점검을 실시한 결 과 웹기반 정보분석시스템 서버와 개인 클라이언트 컴퓨터 와의 데이터 전송속도가 낮은 네트워크상에서는 시스템이 안정하게 데이터 변환을 하는데 문제가 발생하였다. 따라서 기존 로직을 최대한 간소화하여 로드 최적화를 기하였으며 사용이 적은 로직은 최소화하도록 설계하였다.

분석시스템을 활용한 일련의 분석절차는 데이터 변환부 터 클라이언트 PC에 최종 데이터 전송까지 로직내 하나의 오류라도 발생하면 잘못된 결과 데이터가 연산되므로 로직 구성에 매우 신중하여야 하며, 많은 오류 점검과 데이터 전 송 테스트를 통하여 최적의 시스템 환경이 되도록 구성하 는 것이 매우 중요하다. 시스템을 활용한 분석에 앞서 다음 과 같은 작업이 선행되어야 한다. 먼저 연구테마를 선정하 후 연구추진전략을 구축하며, 이에 따라 문헌정보 자료수집

을 위한 구체적인 DB 선정 및 검색을 통하여 가장 효율적 인 모집단을 구성한다. 이때 데이터베이스를 선정할 후 연 구테마에 가장 적절한 색인어 및 주제어의 선별이 필요하 다. 이때 노이즈 정보의 수집을 최소화시킬 수 있도록 전문 가에 의한 색인어 및 주제어 검증이 필요하다. 연구테마의 기술 동향 및 현황을 가장 적절하게 나타낸 데이터베이스 모집단이 구성되었다는 가정하에 일련의 분석 모듈을 통해 분석이 이루어질 수 있다.

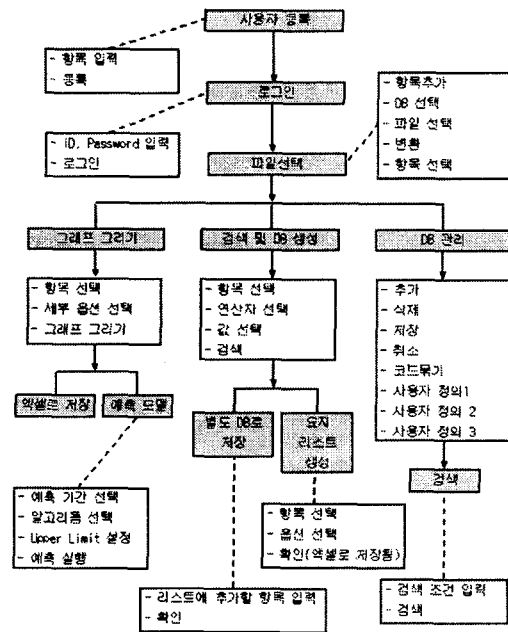


그림 5. 시스템 흐름도

2. 분석 모듈

분석 모듈은 기술수준 분석로직과 기술분야 분석로직으 로 나누어진다. 먼저 COMPENDEX DB를 통해 수집된 정 보는 변환 모듈을 통해 PARADOX 데이터베이스로 생성된 후 기술수준 분석과 기술분야 분석 로직과 연동하여 분석 된 그래프를 나타내 준다. 연도별 논문건수, 연도별 누적 논문 건수, 포트폴리오 맵, 연구성숙도 평가 등을 통해 개 별기술의 개발 동향 및 Life Time 등을 분석하여 현재 기술 수준 분석을 수행하게 된다. 연도별 처리코드 분포, 연도별 분류 코드 분포, 연도별 분류코드 리스트, 연도별 추가된 분류 코드 등을 통해 기술 분야 분석을 하는데 있어 5자리 로 구성된 분류 코드는 400 series의 토목 환경부터 900

Series의 산업공학 분야까지 다양하게 분류되어 개별 기술의 구체적인 기술 분야를 분석해 낼 수 있으며 그래프 분석이 불가능한 경우에는 매트릭스 분석기법을 적용하여 분석 효율을 극대화시켰다.

이외에 추가 그래프 기능을 통하여 기본이 되는 막대 그래프 외에 다양한 그래프 표현이 가능하며, 데이터 엑셀 출력기능을 추가하여 미흡한 그래프기능을 보완하였다. 각각의 분석 그래프내에 구분기능을 추가하여 기본이 되는 연도별 분석외에 저자별, 분류코드별 등의 Z축을 추가하여 3차원 분석이 가능하도록 하였다. 또한, 여기에 랭킹 제한 및 선택 기능, 코드 확인 선택 기능 등을 추가하여 사용자 제한 분석 또한 가능하도록 설계하였다.

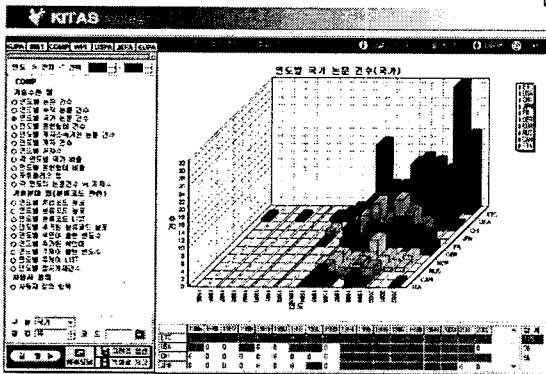


그림 6. 분석모듈 화면

3. 검색 및 DB 생성 모듈

본 모듈은 전체적으로 별도 DB 생성 기능, 유지리스트 작성기능, 복합 연산 검색기능으로 구성되었다. 본 모듈의 기능은 생성된 데이터베이스 모집단의 효율적인 검색과 이를 기반으로 하여 하부 DB를 생성하도록 하는데 있다. 각 DB의 기본 필드와 사용자 필드를 기반으로 하여 검색하도록 하며 검색 효율을 높이기 위해 다양한 연산자를 이용하도록 하였다. 검색된 자료는 새로이 하부 DB로의 생성이 가능하며, 엑셀 출력을 통해 유지 리스트 작성 및 추가 그래프 분석이 가능하도록 설계되었다.

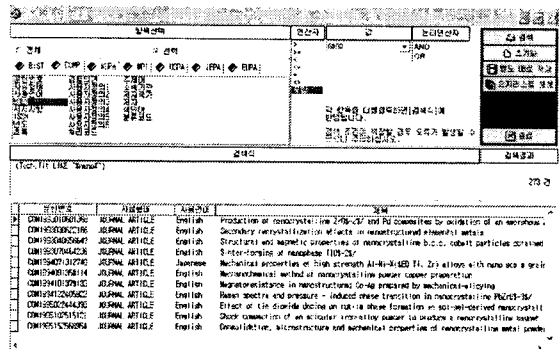


그림 7. 검색 및 DB 생성모듈 화면

4. 콘텐츠 관리 모듈

데이터베이스에서 다운로드 받은 자료를 DB 변환을 통해 변환한 후 그대로 정보 분석에 이용하기에는 많은 문제가 있다. 앞서 언급한 바와 같이 중요한 자료가 색인어, 분류 등의 문자 및 규칙 오류로 인해 잘못 분류될 수도 있으며, 대량의 정보를 연산할 때에는 많은 연산 오류가 발생하여 작게는 모듈내부에서, 크게는 전체 분석시스템의 오류를 야기할 수 있다. 이를 방지하기 위해 DB 관리 모듈을 개발하게 되었으며, 이 모듈은 개별 자료의 추가 삭제 및 수정은 물론 대표명화 작업을 통해 동일한 저자 및 기관의 서로 다른 명칭을 하나로 통일하는 기능 등을 구축하였다.

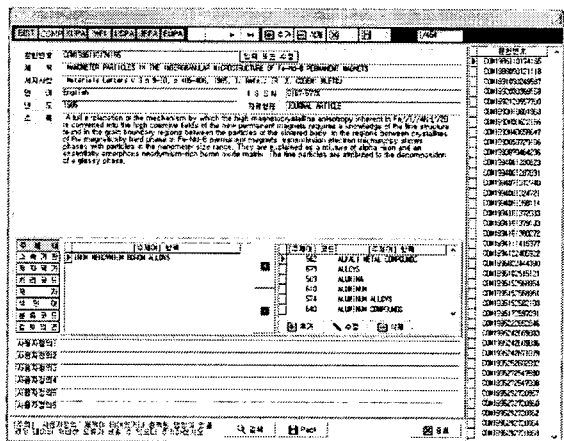


그림 8. 콘텐츠 관리 모듈 화면

또한 DB 관리모듈에는 사용자 정의 기능을 추가하여 전문가가 새롭게 필드를 추가하여 생성할 수 있도록 함으로

써 분석 활용성을 극대화 시켰다. 이때 사용자 정의에 의한 추가 필드값은 초록과 청구 범위 및 분류 코드 등을 전문가가 분석한 후 기입하도록 구성되었다. DB 관리 모듈에는 검색기능이 있어 500건 이상이 되는 모집단에서 사용자가 원하는 항목으로 바로 찾아가 수정·편집·삭제 등이 가능하다. 따라서 검색시 제외시키지 못한 노이즈 데이터를 본 모듈에서 각각 제거할 수 있어 모집단의 주제 정확성을 높일 수 있다.

V. 결론

정보화 사회로 급속히 변모하면서, 정보의 분석 능력이 절대적으로 중요해지고 있다. 기술정보의 체계적인 분석을 통해 기술개발시 방대한 기술정보의 활용성을 높이며, 기술정보의 경제적 가치를 극대화할 수 있는 분석 방법의 정립이 필요하다. 본 연구에서는 전반적으로 이러한 분석 수행할 수 기본적인 인프라를 구축했다고 할 수 있다. 즉 정보 및 DB의 계량학적 접근과 분석틀의 설계 및 그러한 방법을 효율적으로 이용하기 위한 시스템화 등을 전개하였다. 앞으로 이러한 인프라적 연구가 계속 진행되어 정보를 통하여 새로운 정보를 예측할 수 있는 기틀이 마련되어야 할 것이다.

참 고 문 헌

- [1] 한국과학기술정보연구원, "정보검색 이론과 실제", p.456, 1999.
- [2] 김경호, "문헌정보분석을 통한 기술예측 시스템 개발", p.353, 한국과학기술정보연구원, 2001.
- [3] Michael, P. Bigwood, "Analytic tools for the information overload world," Chemtech, pp.12-17, 1997.
- [4] Okubo, Yoshiko, "Bibliometric indicators of research system : Methods and Examples," OECD STI working papers, 1997.
- [5] 유영복, "특허정보에 의한 기술동향분석", 전자공학회지, 제18권, 제3호, pp.34-41, 1991.
- [6] 유사라, "정보학연구와 분석방법론", p.352, 나남출판, 1999.
- [7] 특허청, PATENT MAP, p.206, 1992.

이 호 신(Ho-Shin Lee)

정회원



1991년 2월 : 성균관대학교 금속공학과(학사)

1996년 8월 : 성균관대학교 금속공학과(공학석사)

2001년 3월 : 성균관대학교 신소재공학과(공학박사과정)

1996년 8월 ~ 현재 : 한국과학기술정보연구원, 나노기술정보분석실, 선임연구원

<관심분야> : 나노기술, 응고기술, 표면처리, 정보분석

이 일 형(Il-Hyung Lee)

정회원



1980년 2월 : 한양대학교 전기공학과(학사)

1983년 2월 : 한양대학교 전기공학과 대학원(공학석사)

1995년 3월 : 한양대학교 전기공학과 대학원(공학박사)

1998년 ~ 1999년 : 중국청화대학 Post Doc.

1987년 ~ 현재 : 한국과학기술정보연구원 나노기술정보분석실 선임연구원

<관심분야> : 나노소자, 전력전자, 정보기술, 정보분석

김 경 호(Kyung-Ho Kim)

정회원



1979년 3월 : 서울대학교 화학공학과(학사)

1981년 3월 : 한국과학기술원(KAIST) 화학공학과 대학원(공학석사)

1997년 3월 : 서울시립대 환경공학과 대학원(박사수료)

1984년 ~ 현재 한국과학기술정보연구원 나노기술정보분석실 실장책임연구원

<관심분야> : 나노기술정책, 대기환경, 정보분석, 촉매기술