

사실정보 데이터베이스에 관한 고찰

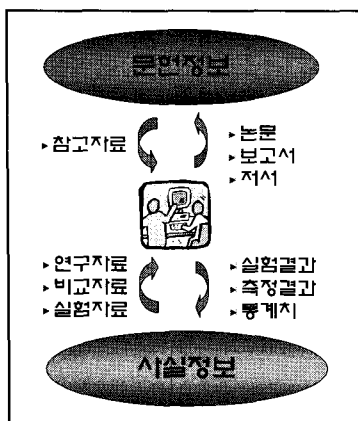
안부영, 이상호 (한국과학기술정보연구원(KISTI))

차 례

1. 서론
2. 사실정보 정의
3. 사실정보 분류방법
4. 사실정보 국내외 현황
5. 결론

1. 서론

현대는 과학기술의 급격한 발전과 디지털혁명으로 개인의 생활양식은 물론 정치, 경제, 사회, 문화, 교육 등의 전 분야를 변화시키는 문명사적 대전환의 시대이고, 정보화 사회를 지나 지식기반 정보사회로 이전하고 있다. 이런 가운데 연구개발을 통해 생산된 과학기술 정보는 데이터베이스 또는 지식베이스로 구축되어 고부가가치의 경제 매체 및 연구개발의 참고자료로 활용되고 있다. 이렇게 과학기술 연구개발에 따라 산출되는 모든 정보를 사실정보(Fact Data)라고 하고, 이런 사실정보를 데이터베이스로 구축한 것을 사실정보 데이터베이스(Factual Database)라고 말한다.



▶▶ 그림 1. 문헌정보 VS. 사실정보

정보의 대표적인 형태인 문헌정보는 연구개발 활동에 간접적인 효과를 갖는데 비해 사실정보는 연구에 직접 활용될 수 있는 정보라고 할 수 있다. 실제로 많은 분야에서 연구자들이 기존의 사실정보를 출발점으로 하여 새

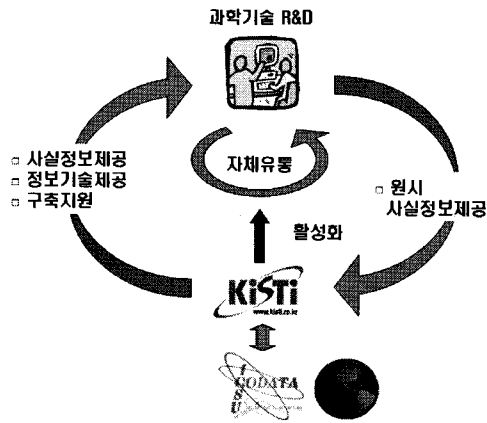
로운 연구를 시작하고 있다. 예를 들면, 물질의 특성 및 표준에 대한 사실정보는 이 물질을 활용하고자 하는 국내외 연구자들에게 없어서는 안 될 필수적인 정보이기에, 데이터베이스로 구축하여 연구를 수행하는데 이용되고 활용되도록 하여야 한다.[그림 1 참조]

사실정보는 있는 그대로의 원시 데이터를 사용자의 요구에 맞게 멀티미디어 형태로 가공한 정보로서 정보기술의 발달로 인한 시대적 변화에 부응하는 정보라고 할 수 있다. 이런 시대에 사실정보 데이터베이스와 같이 전문가들이 구축하고, 그 전문가들이 다시 활용할 수 있도록 서비스가 제공되는 정보는 그 정보요구를 만족할 수 있으며 과학기술의 발전을 위해서도 매우 필요하다고 할 수 있다.

국내에서 사실정보라는 단어가 거론되고 데이터베이스로 구축되어 온라인상에서 정보 서비스 제공이 시작된 지는 20년이 지났다. 그렇지만 사실정보에 대한 전문가들과 일반인들의 인식이 부족한 실정이다. 왜냐하면 사실정보 특성상 과학기술 전문가들에 의해서 생산 및 이용되는 특징이 있는 정보임에도 그 전문가들이 사실정보라는 단어의 의미를 이해하지 못하여서, 본인들이 생산하는 데이터가 사실정보라는 사실을 모르기 때문이다. 사실정보를 직접 생산하는 전문가들도 이러한 이유 일반인들은 더구나 사실정보에 대한 정확한 개념에 관한 인식이 부족한 것은 당연하다 할 수 있다.

따라서 사실정보의 정확한 개념이 무엇이고, 국내외 관련기관은 어디인지, 국내에서는 어떤 종류의 사실정보 데이터베이스가 구축되어 서비스되고 있는지를 정리하여 알릴 필요가 있다고 본다. 이에, 본고에서는 사실정보에 관한 그 동안의 연구를 기반으로 하여 명확한 개념을

재정립하고, 국내의 관련기관의 현황과 국내에서는 유일하게 사실정보 데이터베이스를 구축하여 서비스[그림 2 참조]하고 있는 한국과학기술정보연구원(KISTI)의 사실정보 데이터베이스 구축현황을 소개하고자 한다.



▶▶ 그림 2. KISTI 사실정보 구축 체계

2. 사실정보 정의

사실정보 또는 사실정보 데이터베이스(Factual Database)란 화합물의 구조, 물질의 특성, 유전정보, 생물종 정보 등과 같이 과학기술 분야의 실물에 관한 정보의 데이터베이스를 총칭하는 개념으로 정의할 수 있다.

사실정보 데이터베이스는 일반적으로 전 세계의 여러 과학기술자들의 연구결과로 만들어지는 것이며, 모든 과학자들이 연구개발에 참고할 수 있으므로 국제적으로 공개 및 공유가 이루어져야 한다. 또한, 사실정보 데이터베이스의 구축과 서비스는 상업적인 형태가 되기 힘든 것이 많기 때문에 선진각국에서는 정부의 지원으로 공개된 사실정보 데이터베이스가 지속적으로 생산 서비스되도록 유도하고 있다. 또한, 대부분의 사실정보 데이터베이스는 전 세계 공통의 것이므로 동일한 내용의 것을 각 나라마다 따로 만드는 것은 비용적으로 낭비를 가져온다고 볼 수 있다.

이런 일련의 활동의 전 세계적인 협조 체계를 위하여 과학기술자의 유엔이라 할 수 있는 국제과학기술총연맹(ICSU: International Council of Science Union) 산하 과학기술데이터위원회(CODATA: Committee on Data for Science and Technology)가 1966년에 결성되어 현재까지 활발한 활동을 해오고 있다. 대한민국은 1990년

경부터 참여를 시작하였으며, KISTI 원장이 한국대표로서의 임무를 수행하고 있다.

우리나라는 국가 경제력에 비하여 다른 나라보다 사실정보 데이터베이스 구축과 CODATA 활동이 매우 미약한 형편이며, 이는 과학기술분야에 있어서 우리나라의 국제적인 위상에도 나쁜 영향을 끼치고 있는 것이 사실이다. 우리나라 연구자들은 연구개발을 위하여 사용하는 데이터의 많은 부분을 무료로 공개된 외국의 것을 쓰고 있다. 인터넷을 통한 무료서비스 이외에 책자나 CD-ROM, 유료사이트 등의 비용을 지불하고 사용하는 경우에도 일반적으로 그 데이터 자체에 대한 가격을 지불하는 것이 아니라 콘텐츠에 대한 가격과 우송료 등의 약간의 서비스비용 정도를 지불하는 것이라고 볼 수 있다. 왜냐하면 그 나라 정부의 지원으로 전 세계적인 공개를 원칙으로 만들어진 것이고 그 속에 담긴 데이터를 생산하는 비용과 노력은 매우 크기 때문이다. 이처럼 우리가 외국의 것으로부터 혜택을 받고 있으므로 우리 또한 국가의 경제력과 그 밖의 규모에 알맞은 수준의 국제적 공헌을 하는 것이 당연할 것이다. 비단 이 점에서 뿐만 아니라 국내 과학기술의 발전을 위해서도 국내에서 생산되는 사실정보 데이터베이스의 체계적인 데이터베이스화는 반드시 필요한 일이다.

3. 사실정보 분류방법

3.1 과학기술 데이터 분류(CODATA, 1985)

일반적으로 데이터는 사실적인 것, 숫자, 또는 정보로 분류되고 있으나 CODATA Task Group on the Accessibility and Dissemination에서 고안한 분류 방법에 의하면 과학기술 데이터는 다음의 6가지로 나뉘어진다.(Watson, 1985)

- ① 시간(Time factor)에 의한 분류: 시간의 흐름에 따라 변하거나 신규로 추가되는 데이터
- ② 장소요인(Location factor)에 의한 분류: 장소에 의해 변하거나 신규로 추가되는 데이터
- ③ 발생이나 파생방법(Mode of generation or derivation)에 의한 분류
 - Primary data : 특정 값(Particular quantity)의 측정을 위해 특별하게 고안된 실험에 의해 얻어진 데

이터

- Derived data : 이론적 모델의 도움으로 여러 primary data의 조합에 의해 얻어진 데이터
 - Theoretical data : 이론적 계산이나 예측에 의해 생성된 데이터
- ④ 정량적 값의 성질(Nature of the quantitative value)에 의한 분류
- Determinable data : 어떤 양에 대해 일정한 조건에서는 일정한 값이 얻어진다고 생각될 수 있는 데이터
 - Stochastic data : 어떤 양에 대해 샘플에 따라, 측정 방법에 따라, 같은 조건이라도 다른 값을 가지는 데이터
- ⑤ 표현용어(Terms of expression)에 의한 분류
- Quantitative data: 잘 정의된 정량적 단위로 표현될 수 있는 과학적인 값의 측정치
 - Qualitative data: 과학적 물체, 현상, 성질 등에 대해 정상적인 기술을 포함하는 데이터
- ⑥ 표시방법(Mode of presentation)에 의한 분류 : 이해를 돕기 위한 여러 가지 형태의 표현에 대한 것
- Numerical data(수치 데이터)
 - Graphical data(그래픽 데이터)
 - Symbolic data(기호 데이터)

3.2 데이터 종류별 분류

데이터의 또 다른 분류방법으로는 [표 1]과 같이 CODATA에서 분류한 데이터의 성격에 따른 종류별 분류법으로 사실정보를 크게 Class A, Class B, Class C로 분류하는 것이다(Lide, jr., 1985).

표 1. 데이터 종류별 분류(CODATA)

분류	정의	Data 예
Class A (Hard data)	<ul style="list-style-type: none"> • 잘 정의된 시스템에서 반복 측정에 의하여 획득된 데이터 • 공인기관에 의해 정밀도 및 정확도에 대한 평가를 받은 데이터 	표준 물리상수, 열역학 테이블, 물리·화학 데이터 등
Class B (Semi-hard data)	<ul style="list-style-type: none"> • 관측에 의해 얻어지는 데이터 • 시공간에 따라 변하는 특성을 지닌 데이터 	기상 관측 데이터, 수질 데이터, 생물관련 데이터 등
Class C (Soft data)	<ul style="list-style-type: none"> • 통계 데이터 	인구 통계 데이터, 전염병 데이터 등

Class A(Hard data) 데이터는 잘 정의된 시스템에서 반복 측정하여 얻은 데이터로서 공인된 기관에 의해 정밀, 정확도에 대한 평가를 받은 데이터를 말하며, 물리, 화학 데이터, 수백개의 화학반응에 관한 rate constants, 태양의 자외선에 의한 cross sections for photodissociation 등을 예로 들 수 있다. Class B(Semi-hard data) 데이터는 관측 데이터로서 시간과 공간에 의해 변하는 데이터를 말하며, 지구과학 등의 관측에 의한 데이터, 기상관측 데이터, 생물 관찰 데이터 등을 예로 들 수 있다. Class C(Soft data) 데이터는 통계 데이터를 말하며, 인구통계 데이터, 전염병에 대한 데이터 등을 예로 들 수 있다.

3.3 데이터 형태별 분류

데이터베이스를 분류하는 기준은 매우 다양하지만 형태적인 분류기준으로 텍스트 형태로 제공되는 문헌정보 데이터베이스와 비문헌정보 데이터베이스로 크게 분류할 수 있다. 문헌정보 데이터베이스의 종류로는 학회논문, 학위논문, 연구보고서 등이 있다. 비문헌정보 데이터베이스는 텍스트, 그림, 소리 등으로 구성되는 멀티미디어 데이터베이스와 과학기술과 관련되어 생산·유통되고 있는 사실정보 데이터베이스로 나눌 수 있다(JDBC, 1999).

사실정보를 비문헌정보의 대표적인 형태라고 보았을 때 그 반대 개념인 문헌정보의 특징에 관하여도 알아야 한다. 문헌이란 그 형태와 종류에 상관없이 정보매체에 기록된 정보의 총칭이며, 문헌정보란 지식기반사회에 대처하기 위하여 정보를 분석하고, 이를 체계화하며, 체계화된 정보를 처리하여 정보업무 전반에 필요한 다양한 기술과 이론을 정리한 정보로서, 일반적으로 텍스트 형태로 제공되는 것을 말한다. 문헌정보 또한 컴퓨터와는 분리해서 생각할 수 없기 때문에 문헌정보를 연구하는 문헌정보학은 컴퓨터를 통한 정보 분석·가공에 관한 방법론, 인터넷 등의 가상공간에서 이루어지는 정보의 생산과 활용을 위한 방법론 등의 분야로 그 연구영역을 넓혀가고 있다.

[표 2]에서 보는 바와 같이 사실정보가 존재하는 형태는 수치, 도형, 화상, 영상 등 다양하며, 세계적으로 많이 활용되고 있는 사실정보의 구체적인 예는 [표 3]과 같다.

표 2. 데이터 형태별 분류(일본 데이터베이스백서, 1999)

DB 구분	형태	DB 예
서지(Reference)	문자 정보	서지 정보, 초록
사실(Fact)	수치 정보	통계, 실험수치
	문자 정보	문헌, 기사 전문(full text)
	화상정보 (정지화상)	심전도, 설계도, x-ray 사진
	영상정보 (동화상)	자연 및 실험관찰, 스포츠 기록
	음성 정보	자연음, 기계음, 음악
멀티미디어	혼합된 정보	비디오, cd-rom

표 3. 사실정보 예

- Fundamental physical constants
- Thermodynamic tables
- Properties of engineering materials
- Protein and DNA sequence data
- Microbial strain data
- Satellite data
- 약 성분 data
- 천문기상 data
- Biodiversity data

4. 사실정보 국내외 현황

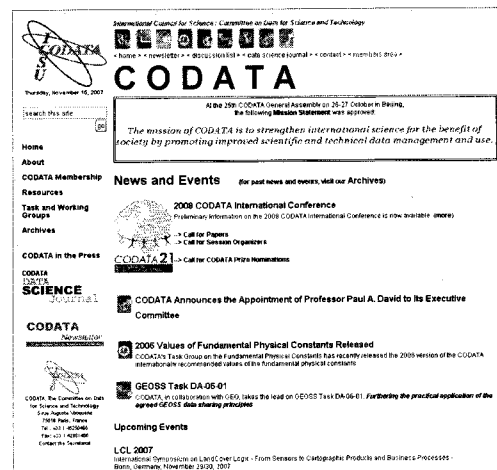
사실정보는 방대한 양으로 생산되고 있는 실험관측 데이터, 정량적 정보 데이터 등의 수치데이터를 포함하는 실질적인 과학정보(Factual Scientific Information)를 포함한다. 이러한 데이터의 양이 많아질수록 데이터베이스화하고 웹을 통해 검색 가능한 시스템을 개발하는 것이 요구되어 진다. 또한 과학기술 정보의 축적과 공유는 비영리적인 행위이고 무상 제공을 목표로 하므로 공공기관이 아니면 할 수 없는 일이라고 할 수 있다. 또한 구축된 국내 사실정보 데이터베이스는 해외와의 과학기술 데이터교환 프로그램에 적극적으로 참여하는 데 중요한 자원이 될 뿐 아니라 지식기반사회의 훌륭한 고유 자산이 된다. 이런 취지로 한국과학기술정보연구원(KISTI)에서는 1994년부터 지금까지 사실정보를 발굴, 수집, 분석, 구축하여 서비스를 제공하고 있다.

다음부터는 사실정보 국외 현황 및 KISTI에서 구축하여 서비스하고 있는 과학기술 사실정보 데이터베이스에 관하여 살펴보도록 하겠다.

4.1 국외 현황

4.1.1 CODATA(Committee on DATA for science and technology)

사실정보 관련 대표적인 국제기구로는 과학기술데이터위원회(CODATA: Committee on DATA for science and technology)가 있다. CODATA는 국제과학기술총연맹(ICSU: International Council of Science Union)의 특별위원회로 1966년에 설립되었으며 첫 총회는 독일의 Frankfurt에서 열렸다. CODATA의 설립목적은 각국에서 급격히 생성되고 있는 방대한 양의 사실정보의 질을 평가하고 데이터베이스화하여 국제적 상호공유 및 교환을 촉진하고 사실정보 데이터베이스 구축에 관한 국제적인 협력 체제를 유지하는 것이다. CODATA의 활동 목표는 과학기술 발전에 따라 발생하는 지식을 유용한 기술로 전이시키는데 필요한 신뢰할 수 있는 데이터를 제공하는데 있다.

▶▶ 그림 3. CODATA 웹사이트 (<http://www.codata.org>)

CODATA의 활동 목표를 좀 더 자세히 살펴보면, 첫째, 데이터 제작자와 사용자 사이의 협력 증진, 둘째, 데이터 수집, 배포, 평가를 장려하고 데이터의 품질을 조정, 관리하여 이 분야의 국제협력촉진, 셋째, 여러 학문간의 데이터 신뢰도를 제고하고, 데이터를 이용의 편리한 방법 모색, 넷째, 중요한 과학 데이터의 수집 및 수집 장려, 다섯째, 데이터 수집 및 편집의 방법을 개선하기 위한 방법 연구, 여섯째, 데이터베이스 작성이나 데이터베이스센터를 설립하기 위한 타당성연구, 일곱째, 과학자들에게 이러한 활동의 중요성을 인식시키고 참여 장려 등으로 정리할 수 있다.

초기의 CODATA는 단순히 데이터 수집 및 배포에 관심이 있었지만 최근에는 인터넷의 발달과 더불어 데이터의 디지털화와 웹을 통한 공유를 강조하고 있다. CODATA에서 수행하고 있는 주요활동은 다음과 같다.

- 국제적으로 계속 이용할만한 기본 데이터 Set 작성: 과학기술의 중요한 분야에 믿을만한 데이터 Set을 만든다. 그 예로 Task Group을 결성하여 기본 데이터 set을 만든다. 1973년 기본 상수 값(fundamental constants)을 거의 모든 학술기관의 동의를 얻어 조정하였고 신규 조정안이 1984년에 발표되었다.
- 데이터베이스의 상호 이용을 위한 데이터베이스의 표준화 작업 정립: 데이터베이스를 서로 교환하여 쓸 수 있도록 데이터베이스와 DBMS의 표준화 작업을 수행한다.
- 믿을만한 데이터 소스에 대한 정보 제공: CODATA는 여러 과학 분야에서 믿을만한 데이터를 얻기 위해 활용 가능한 데이터 자원의 종류 및 소스에 대한 정보를 제공하고 있으며, 이를 위해 분야별로 Task Group을 운영하고 있다.
- 교육과 훈련: UNESCO와 다른 곳에서 보조를 받아 CODATA가 데이터 관리하는 훈련코스(Training Courses in Data Handling)를 운영하고 있다.
- 여러 국가가 참여한 프로젝트에 협조: Data bank on hybridomas and monoclonal antibodies, Network of Thermodynamic data centers 등의 프로젝트를 협력 운영하고 있다.
- Conference 개최: 과학기술 정보관련 컨퍼런스를 짝수년도에 회원국이 돌아가면서 개최하고 있다. 2008년에는 10월 5일-8일까지 우크라이나에서 컨퍼런스를 개최한다.

4.1.2 JST(Japanese Science and Technology Cooperation)

과학기술진흥사업단(JST)은 과학기술정보 유통업무를 수행하던 일본과학기술정보센터(JICST-1957년 8월 설립)와 기초연구, 신기술개발, 연구교류 등의 업무를 수행하던 신기술사업단(1961년 7월 설립)이 통합해서 양기관이 수행하던 사업을 승계·개발·추진을 목적으로

1996년 10월에 설립된 기관이다. 일본의 과학기술정보에 관한 중추기관이며 과학기술 정보유통, 기초연구, 기술이전, 연구교류, 연구지원, 과학기술 이해 증진, 과학기술의 진흥에 기여하는 기반 업무 등을 수행하고 있다. 물질, 재료, 생명공학 등의 고기능 데이터베이스를 개발 및 운영 중이며, 사실정보 성격을 가진 데이터베이스는 [표 4]와 같다.

표 4. JST 제공 사실정보 데이터베이스

<ul style="list-style-type: none"> ○ FACTrio: Kelvin, Dalton, Angstrom ○ Alloy Database ○ PolyInfo ○ Bioinformatics DB
--

4.1.3 FIZ Karlsruhe(Fach informationszentrum Karlsruhe)

독일의 FIZ Karlsruhe는 학계와 산업계의 R&D 정보를 제공하는 기관이다. 미국, 일본과 함께 STN 서비스를 하고 있으며 과학기술에 관련된 정보뿐만 아니라 비즈니스와 관련된 정보도 공개 서비스를 하고 있다. 유럽의 STN 서비스센터로서 특허를 포함한 과학기술에 관한 거의 모든 분야의 200개가 넘는 데이터베이스를 제공하고 있다.[표 5 참조]

표 5. FIZ Karlsruhe 제공 데이터베이스

<ul style="list-style-type: none"> ○ ENTEC(formerly ENERGIE) ○ ENERGY(DOE Energy file) ○ INIS(Nuclear research and technology) ○ ICSD(Crystallography) ○ INSPEC(Physics subset)
--

4.1.4 NIST(National Institute for Standards and Technology)

NIST는 1901년 미국 의회에 의해 NBS(National Bureau of Standards)로 설립되었으며, 1988년 NIST(National Institute for Standards and Technology)로 개명되었다. 미 상무성(Department of Commerce) 소속의 정부기관으로 미국의 전반적인 산업 경제를 향상시키기 위해 업계와 협력하여 기술, 기준, 표준을 개발하고 이를 적용하고 지원하는 것을 주목적으로 하고 있다.

NIST의 참조표준정보에 대한 활동은 40년 이상 지속되고 있다. 1963년에 미 연방정부에서 국가 참조표준 데이터 시스템에 대한 연방 정책(National Policy on

National Standard Reference Data System)을 수립하였으며, 1968년 미 의회에서 공법 90-396으로 승인되면서 참조표준에 대한 활동을 본격적으로 시작하였다.

현재는 참조표준정보 프로그램(Standard Reference Data Program)으로 이름이 바뀌었으며 이 프로그램을 운영하는 목적은 신뢰성 있는 평가된 수치 데이터의 보급이다. NIST에서 서비스하는 데이터베이스는 67개 정도이며 그중 대표적인 데이터베이스는 [표 6]과 같다.

표 6. NIST 제공 데이터베이스

- Analytic Chemisrty DB
- Atomic and Molecular Physics DB
- Biotechnology
- Chemical and Crystal Structure
- Chemical Kinetics
- Industrial Fluids and Chemical Engineering
- Thermochemical DB
- Materials Properties DB
- Surface data
- Property DB

4.2 국내 현황(KISTI 현황)

국내에는 사실정보로 분류할 수 있는 데이터베이스들이 다수 존재하나, 사실정보 데이터베이스에 관하여는 KISTI가 시스템공학연구소 시절 국제기구인 CODATA의 회원기관으로 활동을 하면서 사실정보 데이터베이스 구축사업을 수행한 것이 그 시초라고 할 수 있다.



▶▶ 그림 4. KISTI 사실정보 데이터베이스 웹사이트 (<http://fact.kisti.re.kr>)

KISTI는 국가 과학기술정보 분야의 전문연구기관으로 과학기술 및 이와 관련된 산업에 관한 정보를 종합적으로 수집·분석·관리하고 정보유통에 관한 기술·정책·표준화 등을 전문적으로 조사·연구하며 연구개발 인프라를 체계적으로 구축·운영함으로써 국가 과학기술 및 산업발전에 이바지함을 목적으로 설립된 과학기술 정보 유통기관이다.

KISTI에서 분야별 전문가의 도움을 받아 구축하는 사실정보 데이터베이스의 구축 범위는 크게 네 가지로 구분할 수 있다. 첫째로 산·학·연 연구자들을 위한 연구개발 지원 데이터베이스로써 국내에서 발견되고 분류된 생물 관련 정보와 재료 물성 데이터, 실험데이터 등이 있다. 둘째로 인기는 없지만 국가적으로 보존가치가 있는 국내 고유 정보를 데이터베이스로 축적하여 National Archives Center의 역할을 수행하고 있다. 세 번째로는 일반인이나 분야가 다른 전문가들에게 다른 분야의 정보를 전할 수 있는 과학기술 관련 데이터베이스가 있다. 네 번째로는 학습 지원 데이터베이스로써 초·중·고등학생의 학습 콘텐츠로 활용 가능하고, 교사들의 학습 보조 콘텐츠로 활용 가능한 Virtual Science Museum 형태의 멀티미디어 정보가 있다. 다음은 KISTI 과학기술 사실정보 데이터베이스 웹사이트(<http://fact.kisti.re.kr>)에서 서비스되고 있는 사실정보 데이터베이스에 관하여 간략하게 소개하도록 하겠다.

4.2.1 화합물 DB(ChemDB)

신약설계를 위한 화합물 데이터베이스 ChemDB는 한국과학기술정보연구원에서 주관한 “과학기술 지식정보 시스템 구축” 사업으로 2000년부터 2006년까지 사단법인 분자설계연구소(BMD)의 기술 지원을 통해 구축된 화학 및 의약, 생물 분야에서 활용될 수 있는 전문 데이터베이스이다. 본 데이터베이스는 현재 유료화 시범사업으로 지정되어 일부 데이터베이스 검색 결과를 sd 파일로 다운로드 받는 것을 유료로 운영하고 있다.

본 데이터베이스에서는 화합물의 2차원 분자구조와 3차원 분자구조, 화합물의 물성 및 여러 가지 표현자 값 (Topological Descriptors, calculated logP, calculated logS, H bond acceptors, H bond donors, Topological Polar Surface Area, Moment of inertia, Surface area, Volume, Dipole Moment, HOMO energy,

LUMO energy, Polarizability) 등 화합물에 대한 다양한 정보를 포함하고 있으며 데이터베이스에 포함된 다양한 정보를 유용하게 활용할 수 있도록 여러 가지 구조 검색 및 검색 결과 필터 기능을 제공한다.

4.2.2 무기결정구조 DB(ICSD)

무기결정구조 DB는 21세기 고부가가치 산업중 하나인 재료분야 연구를 지원하기 위하여 독일 Fiz-Karlsruhe의 CD-ROM 버전인 무기결정구조정보(ICSD)를 국내 라이선스 허락을 얻어 웹 버전으로 다시 개발한 것이다. 본 데이터베이스는 1915년부터 자료가 수집되기 시작하여 현재까지 약 90,000건의 데이터가 수록되어 있으며 매년 2,000건 이상의 데이터가 업데이트되고 있다. 무기결정구조 DB에서는 결정구조 및 서지 정보, XRD 패턴정보, 3차원 결정구조 정보 등을 웹기반 검색 시스템과 XRD 패턴 검색 기능, 3차원 결정구조 표현 등의 검색 방법을 제공한다. 좀 더 자세한 검색방법은 다음과 같다.

Chemistry 검색은 화합물에 포함된 원소를 선택하여 입력할 수 있게 주기율표 형식의 사용자 인터페이스를 제공하고 있으며, Crystal Data 검색은 단위 셀의 cell parameter(a, b, c, α , β , γ) 및 cell volume(V) 값의 범위, 공간군 번호를 지정하여 검색하고자 할 경우 사용자가 값을 직접 입력하는 방식과 버튼을 클릭하여 보여지는 Space Group을 선택하여 검색하는 방식이 있다.

Reduced Cell 검색은 결정축계(P, C, I, F, A, B, Rr, Rh)를 선택하거나, 셀 파라미터를 직접 입력, Tolerance도 원하는 값으로 입력하여 검색이 가능하다. Symmetry 검색은 결정계(Triclinic, Monoclinic, Orthorhombic, Tetragonal, Trigonal/Rhombohedral, Hexagonal, Cubic)와 결정축계(P, A, B, C, F, I, R)를 선택하거나, Space group symbol, Wyckoff Sequence, Laue/Crystal Class, Pearson Symbol 중 하나를 지정, 원하는 공간군을 선택하여 검색할 수 있으며 Reference 검색은 논문, 저자, 출처 등의 서지 정보와 광물질에 대한 검색조건을 입력하면 된다.

마지막으로 XRD Pattern 검색은 XRD 패턴 주요 피크의 2-theta를 5개까지 사용자가 입력하여 패턴 검색을 할 수 있으며 X-선 회절장치마다 교정 오차를 고려하여 사용자가 입력한 2-theta값에 대하여 ± 0.5 내의 피크를

포함시키도록 하였다.

4.2.3 플라즈마 물성 DB

플라즈마는 물질의 4번째 상태로 불리우는 고에너지 하전입자 및 중성 입자들의 집합체이다. 플라즈마가 이렇듯 이온화에 의해 발생된 고에너지의 하전입자를 포함하고 있기 때문에 선택적으로 조정할 수 있는 능력, 다른 물질과 활발히 반응하여 대상물질의 상태를 물리, 화학적으로 변화 시키는 능력 및 활성도가 낮은 중성종을 자극하여 활성도를 향상시키는 능력 등의 유일한 특성을 지니게 된다. 이러한 특성으로 인해 핵융합, 반도체 제작, 표면 처리, 폐기물 처리, 디스플레이 장치 등의 다양한 분야에 응용되고 있다.

플라즈마 물성 DB에서는 플라즈마에 의해 촉진되는 화학반응, 플라즈마와 상호작용 물질의 특성, 그리고 플라즈마를 구성하는 분자 및 원자의 구조, 이온화 에너지, 충돌단면적, 플라즈마의 확산계수, 방전전압, 스펙트럼 등의 각종 데이터들을 그래프 또는 수치 형태로 제공하고 있다.

4.2.4 인체 DB

인체 DB는 한국인의 인체영상과 디지털 코리아 등 2개의 데이터베이스가 구축되어 서비스되고 있다. 먼저 한국인의 인체영상(<http://vkh3.kisti.re.kr>)은 한국인 남성 인체에 관한 2차원 및 3차원 정밀영상 DB(Visible Korean Human)를 구축하였다. 본 DB는 기준받은 시신을 0.2mm 간격으로 절단하여 촬영한 영상 데이터를 기반으로 다양한 2차원 영상 브라우징 및 3차원 인체 영상 렌더링 SW를 개발하여 제공하고 있다. 미국 국립보건원(NIH)에서 제공하는 인체영상 데이터도 있지만 우리 것만큼 정교하지 못하여 독일, 미국 등 다른 나라에서도 우리나라의 인체영상 데이터를 제공받아 활용하고 있다.

디지털코리아(<http://digitalman.kisti.re.kr>)은 남녀 각 50 표본을 1mm 간격으로 CT로 촬영한 후 개별 뼈 모델링 및 평균화 골격 모델을 완성하였으며 피부/치아/손/발의 정밀도 향상 및 평균 얼굴 또한 완성하였고 머리/척추/골반/넙다리 물성정보 및 단일뼈 하중정보도 구축하여 서비스하고 있다. 디지털코리아는 산업자원부에서 추진하는 사이즈 코리아 사업에서 제공하는 단순 외

형 치수뿐만 아니라 인체 내부의 정보를 추가적으로 제공함으로써 의료, 바이오, 스포츠 과학 등의 분야에 효과적으로 활용이 가능하다. 그리고 한국인의 인체 정보는 곧 동양인의 인체 정보임으로 국내 기업의 동남아 시장에서 경쟁력 향상에 기여할 수 있을 것으로 기대된다.

4.2.5 생물다양성(Biodiversity) DB

생물다양성은 육상 생태계, 해양과 기타 수생 생태계와 이들의 복합 생태계를 포함하는 모든 원천에서 발생한 생물체의 다양성을 말하며, 종내·종간 및 생태계의 다양성을 포함한다. 즉, 생물다양성이란 생물종(Species)의 다양성, 생물이 서식하는 생태계(Ecosystem)의 다양성, 생물이 지닌 유전자(Gene)의 다양성을 총체적으로 지칭하는 말이다.

이러한 생물다양성 정보를 좀 더 효율적으로 활용하기 위한 생물다양성정보 데이터베이스 구축사업은 국가의 발전을 위해 정보의 인프라를 구축한다는 데 의의가 있다. 생물다양성 정보의 축적과 공유는 비영리적인 행위이고 무상제공을 목표로 하고 있다. KISTI가 중심이 되어 구축한 외부전문가와 구축한 국내 고유의 생물다양성 데이터베이스는 16종정도이며[표 7 참조] 국제 데이터교환 프로그램에 적극적으로 참여하는 데 중요한 자원이 될 뿐 아니라 지식기반사회의 훌륭한 고유자산으로 그 가치가 더하고 있다.

표 7. KISTI 생물다양성 DB 현황

데이터베이스명	구축 건수	정보제공자
국내 담수어류 DB	132종	전북대 김익수교수
곤충(나비/벌목) DB	1,800종	강원대 박규택교수
곤충(벌목:뱀시벌과) DB	300종	영남대 이종욱교수
육상절지동물 문헌 DB	10,296건	충북대 조수원교수
국내 조류분포 DB	375종	경희대 유정철교수
환형동물(지렁이목) DB	40종	전북대 김태홍교수
해충(진드기목) DB	80종	전북대 이원구교수
동물결명 DB	200건	전남대 박남용교수
민속/특산식물 DB	1,000종	전북대 김무열교수
자생식물 종자 DB	600종	동신대 장홍기교수
수생관속식물 DB	70종	아주대 최홍근교수
식물바이러스 DB	100종	영남대 장무웅교수
국내 고등균류 DB	1,250종	우석대 조덕현교수
가상화석박물관	1,050점	호서대 홍성수교수
가상패류박물관	950점	호서대 홍성수교수

4.2.6 생명정보(Bioinformatics) DB

생명과학 기술이 발전함에 따라 연구개발이 급증하면서 실험방법이 발달하게 되었고, 그 결과 산출되는 생명정보 데이터의 양은 기하급수적으로 증가하고 있다. 이런 방대한 양의 생명정보 데이터를 분석하고 분석된 데이터에서 유용한 정보를 얻어내는 생명정보학(Bioinformatics)이 등장하게 되었다. 생명정보학은 생물학, 화학, 물리학, 수학, 의학, 약학 등을 연구하는데 있어 컴퓨터학을 접목시킨 새로운 복합학문 분야로써 대량의 생명정보 데이터베이스, 이 데이터베이스를 시뮬레이션 할 수 있는 분석도구(SW)와 컴퓨터 그래픽 등과 같은 정보기술을 이용하여 대량의 생명과학 데이터를 축적하고 분석하는 계산 생물학(Computational Biology)이라고 말할 수 있다. 물론 이 모든 것을 운용할 수 있는 고성능 시스템(HW)과 초고속 네트워크 환경은 필수적이다.

이에 KISTI에서는 생명과학 연구개발 지원을 위하여 'IT기반 생명정보 인프라 구축'이라는 큰 목적을 가지고, 연구자들이 많이 사용하는 데이터베이스와 분석도구를 수집하여 서비스하고 있다.[표 8 참조]

표 8. KISTI 생명정보 DB 현황(2007년 9월 현재)

데이터베이스명	구축건수	비고
GenBank	76,146,236	release 161
REBASE	4,712	release 607
Ensembl	156,500,000	313Gb(1건=2Kb)
PDB	45,506	2007-09-01
PIR	285,376	
Swiss-Prot	227,883	release 54.1
CATH	48,391	
MHCBN	78,000	
DIP	48,902	
BIND	80,993	
GeneCards	48,393	release 2.36
SCOP	105, 529	release 1.71
Pfam	9,318	release 22.0
InterProScan	4,949,164	release 16.0
2Dgel vips	3,052,919	

4.2.7 천문 DB

천문 DB는 우리나라 고유의 표준연력과 고천문기기 등 2개가 구축되어 서비스되고 있다.

표준연력 DB(<http://manse.kisti.re.kr/>)는 과거 천문현상의 재검토를 바탕으로 역사적 사실을 조사할 수 있

도록 구축한 것이다. 본 데이터베이스는 우리나라의 역사, 서지학 연구에 기본 자료를 제공하며 미래의 계획을 설계하고 천체역학을 이용한 미래의 천체위치 계산이 가능하다. 앞으로 우주 개발 연구 분야에서 인공위성이나 천체의 궤도 계산 분야에도 활용 가능하리라고 본다.

고천문기기 DB(<http://anastro.kisti.re.kr/>)는 천문에 대한 각종 유산을 사상, 역법, 의기, 역사, 서적, 학자별로 분류하고 수집·분석·연구하여 데이터베이스로 구축한 것이다. 우리나라 고유의 천문 자료들을 3D, Animation, 삽화, 사진 등의 Multimedia화를 통하여 정보통신 산업 시대에 부합하는 자료로 재구성하였으며, 이런 데이터베이스화를 통하여 고천문 분야의 효율성을 증대시킬 수 있으리라 판단된다.

4.2.8 가상과학관 분야

기존의 사실정보 데이터베이스를 재가공하여 가상과학관을 구축하여 일반인과 학생을 위한 서비스를 제공하고 있다. 현재 구축되어 운영되는 가상과학관은 KISTI 가상과학박물관, 화석박물관, 패류박물관 등 3개이다.

KISTI 가상과학박물관(<http://vsm.kisti.re.kr>)은 가상의 공간에서 전문적인 과학 기술 관련 정보를 쉽게 접할 수 있도록 기존 사실정보 데이터베이스를 이용하여 멀티미디어 3차원 공간을 구성하였다. 한국 고유의 생물 다양성관, 패류관, 화석관, 천문우주관, 농업전시관 등 5개 전시관과 16개 전시실로 구성되어 있으며, 전문가뿐만 아니라 학생과 일반인들에게 친숙한 인터페이스 구성으로 교육적인 효과를 더할 수 있게 구현하였다.

화석박물관(<http://museumfs.kisti.re.kr>)은 850점의 화석과 공룡에 대한 데이터 수집, 조사, 분류 및 화상, 음향, 동화상 등 1,050건 이상의 데이터를 구축하였다. 구축된 사이트는 현실감을 위해 정지화상과 문자 데이터뿐만 아니라 동화상 등의 멀티미디어 데이터를 포함하여 시각적 검색 기법으로 접근이 가능하도록 기능을 구현하였다.

가상패류박물관(<http://museumfs.kisti.re.kr>)은 950점의 패류에 대한 데이터수집, 조사 분류 및 화상, 음향, 동화상 등의 데이터를 입력하였으며, 공간데이터를 저장, 분석, 관리할 수 있는 패류정보검색시스템 설계하여 구현하였다.

5. 결론

사실정보는 실험수치, 실측정보, 특정분야 내지 특정 대상에 관한 실질적인 정보를 말한다. 문헌정보는 연구 개발 활동에 간접적인 효과를 갖는데 비해 사실정보는 연구에 직접 활용될 수 있는 정보라고 할 수 있다. 실제로 많은 분야에서 연구자들이 기존의 사실정보를 출발점으로 하여 새로운 연구를 시작하게 되며, 연구 분야뿐만 아니라, 산업 분야에서도 사실정보는 중요하게 활용될 수 있다.

이에 본 고에서는 사실정보에 관한 그 동안의 연구현황을 토대로 여러 가지 분류기준을 조사하고, 명확한 개념을 정립하였다. 또한 국외의 현황을 알아보기 위하여 국제기구인 국제과학기술총연맹(ICSU: International Council of Science Union) 산하 과학기술데이터위원회(CODATA: Committee on Data for Science and Technology)의 활동 및 NIST 등의 기관들과 그 기관에서 생산되고 서비스되는 데이터베이스에 관하여 조사하였다. 국내현황을 알아보기 위해서는 국내에서는 유일하게 사실정보 데이터베이스를 구축하여 서비스하고 있는 KISTI의 사실정보 데이터베이스 구축현황을 조사하여 요약하였다.

국내에서는 과학기술 전문정보인 사실정보에 대한 인식이 많이 부족한 상태이다. 그러므로 사실정보 데이터베이스의 구축 및 활용을 위한 노력은 여전히 필요하며, 과학기술 전문정보를 생산하는 산·학·연 전문가 발굴 및 다양한 멀티미디어 정보 관련 기술개발을 통한 사실정보의 적극적인 서비스 또한 필요한 실정이다.

지식기반 정보사회에서 실현 가능한 멀티미디어 기술 등 각종 적합한 기술이 집적되어 사용자 중심으로 제공되는 사실정보 데이터베이스는 사회 과학기술 간접자본으로서 높은 가치를 지닐 것으로 예상되므로 국가 차원에서 국내 사실정보의 발굴 및 데이터베이스 구축과 더불어 서로 다른 형식으로 산재해 있는 사실정보를 통합 서비스할 수 있는 표준 체제를 확립하여야 할 것이다. 또한 국내의 과학기술 전문정보인 사실정보의 보호 및 보급을 위해서도 노력을 기울여야 할 것으로 사료된다.

참고문헌

- [1] Japan Database White Paper, JDPC(일본 데이터베이스진흥센터), 1999년
- [2] 안부영. 2002. KISTI 생물다양성 DB 구축현황. 『지식정보 인프라지』, 통권 10호: 26-49.
- [3] 안부영, 김태희, 김선호. 메타 사이버 과학박물관의 발전 방향과 그 현황, 제2회 디지털도서관 Conference 논문집: 71-75
- [4] KISTI 사실정보 데이터베이스 홈페이지. [인용 2007년 11월 14일] <<http://fact.kisti.re.kr>>
- [5] ICSU CODATA 홈페이지. [인용 2007년 11월 14일]. <<http://www.codata.org>>

저자소개

● 안 부 영(Bu-Young Ahn)

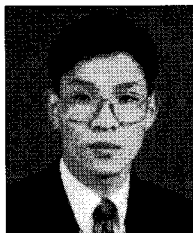
정회원



- 2003년 8월 : 공주대학교 교육정보대학원(교육정보석사)
 - 2007년 2월 : 충남대학교 문헌정보학과(박사 수료)
 - 1982년 3월 ~ 현재 : 한국과학기술정보연구원(KISTI) 선임기술원
- <관심분야> : 사실정보, 메타데이터, 웹 2.0

● 이 상 호(Sang-Ho Lee)

정회원



- 1982년 2월 : 충북대학교 공업화학과(공학사)
- 1984년 2월 : 충북대학교 대학원 공업화학과(공학석사)
- 1993년 3월 : 동경농공대학 유기재료공학과(공학박사)
- 1983년 7월 ~ 현재 : 한국과학기술정보연구원(KISTI) 책임연구원

<관심분야> : 사실정보, 데이터베이스, 인체정보