

국가수준의 과학데이터 시각화 지원체계에 관한 연구

A Study on the Supporting System for Scientific Data Visualization at the National Level

박 동 진* · 채 균 식** · 류 범 중*** · 이 상 태****

Dong-Jin Park · Kyun-Shik Chae · Beom-Jong Ryu · Sang-Tae Lee

차 례

- | | |
|--------------------------------|---------------------|
| 1. 서 론 | 4. 연구개발지원 정책방향 및 과제 |
| 2. 과학데이터 시각화 연구동향 | 5. 결 론 |
| 3. 과학데이터 시각화 지원을 위한 주요영역 설정 | · 참고문헌 |

초 록

전통적으로 과학데이터 시각화는 데이터 분석과 관련이 깊은 연구 수행과정 중의 하나로 간주되어 왔다. 그러나 최근 들어서 과학데이터 시각화의 연구영역 자체가 독립적인 학문분야로서 자리를 잡기 시작하면서 시각화의 기술 및 방법들이 점차 고도화되고 복잡해지고 있다. 그러나 과학데이터 시각화 연구가 개인 및 그룹의 연구자들 스스로의 노력으로 시각화 문제를 해결하는 데 있어서 한계가 있다. 따라서 전문집단에 의한 체계적인 지원의 필요성이 대두되고 있으며, 본 연구는 거시적 관점의 국가수준에서 어떻게 과학데이터 시각화를 체계적으로 지원할 것인가를 정책적인 시각에서 분석하였다. 구체적으로 기존 문헌의 조사를 통하여 주요 성공요인들을 파악하였으며, 집중적으로 투자해야 할 연구의 영역을 설정하였다. 다음으로 국가수준의 시각화 연구지원의 전략적인 방향과 전략적 과제를 제안하였다.

키 워 드

과학데이터, 시각화, 참조표준, e-science

- * 공주대학교 산업시스템공학과 교수
(Professor, Dept. of Industrial and Systems Engineering, Kongju National Univ., mispdj@kongju.ac.kr)
 - ** 한국표준과학연구원 국가참조표준센터 책임기술원(교신저자)
(Corresponding Author, Senior Technical Engineer, National Center for Standard Reference Data, KRISS, cks@kriss.re.kr)
 - *** 한국과학기술정보연구원 지식기반실 책임연구원
(Senior Researcher, Dept. of Knowledge Resources, KISTI, ybj@kisti.re.kr)
 - **** 한국표준과학연구원 국가참조표준센터 책임기술원(교신저자)
(Corresponding Author, Senior Technical Engineer, National Center for Standard Reference Data, KRISS, stlee@kriss.re.kr)
- 논문접수일자: 2011년 3월 21일
• 최종심사(수정)일자: 2011년 4월 7일
• 게재확정일자: 2011년 4월 13일

ABSTRACT

Conventionally, scientific data visualization is thought of as one of activities performed by scientists during the scientific data analysis. However, recently, there exists a set of research papers which count scientific data visualization as an independent research area. They show the research subjects for studying the scientific data visualization technology and methods. In case, a scientist or group of scientists can not solve their own visualization problem due to the unskillfulness and inexperience on using visualization tool. Therefore, it needs to help them by the systematic way for solving the problem. In this study, we analyze and propose the national level scientific visualization support system for scientists. In particular, we first analyze the existing papers and find out the critical success factors. Then, by integrating the findings of the analysis, we propose the research areas which need to be focused, and the strategic direction and specific research topics for scientific data visualization support system in national level.

KEYWORDS

Scientific Data, Visualization, Standard Reference Data, e-science

1. 서론

최근 들어서 과학데이터 시각화 연구영역 자체가 하나의 학문분야로서 자리를 잡기 시작하였다. 시각화는 다른 학문분야를 보완할 뿐 아니라 발전을 촉진한다. 즉 수학과 통계학의 지식들이 전통적인 과학, 경제학, 의학, 사회학 등 여러 학문에서 필요로 하듯이, 이제 시각화도 다양한 학문영역의 연구자들에게 중요한 연구 활동이다. 통계학처럼 시각화도 계량적 그리고 정성적으로 명확한 특성을 전달하면서 정보의 분석과 이해를 돕는 것이다(Johnson et al, 2006).

시각화는 연구과정에 있어서 새로운 사실을 발견하고 연구자의 통찰력을 향상시키기 위하

여 이미지네이션, 컴퓨터 툴, 상호대화용 인터페이스 등을 적절하게 조합하여 사용자의 이해를 향상시키는 것에 목적이 있다고 할 수 있다. 구체적으로 시각화는 연구주제에 관해서 가설을 설정하는 데 도움을 주며, 가설을 평가하고 탐색(exploration)하는 데 도움을 주기도 하며, 새로운 사실의 발견과 함께 과거에 사실로 인정되어 오던 것을 부정하고 새로운 가설을 생산하는 데에도 도움을 준다.

적용 분야를 보면 시각화는 생리학 분야에서 DNA로부터 인체의 전체 조직까지의 복잡한 현상을 이해하는 데 있어서 근본적인 기술을 제공한다. 또한 수세기에 걸친 기후의 변화, 제트날개에 미치는 공기의 흐름을 보여주는 복잡한 현상도 보여주기도 한다. 그리고 시각

화 기술은 데이터스트림을 신속하고 경제적으로 줄여주거나 재구성하게 한다. 이러한 기술의 응용으로 보건분야에서는 지역 혹은 국가적 수준에서 전염병의 확산을 추적하기 위하여 발생하는 대용량 데이터에서 필요한 정보를 파악할 수 있게 한다.

최근 들어서 시각화의 기술 및 방법들이 점차 고도화되고 복잡해지고 있다. 그래서 과학 데이터 시각화 문제를 개인 및 그룹의 연구자들 스스로의 노력으로 해결하는 데 있어서 한계가 있다.

따라서 전문집단에 의한 체계적인 지원의 필요성이 대두되고 있다. 국가수준에서 시각화를 지원할 수 있는 체제의 구축이 필요한데 국내에서는 아직 이와 관련한 선행연구는 전무한 실정이다. 본 연구의 목적은 국가수준에서 시각화를 지원할 수 있는 시스템 체제구축의 주요 성공요인과 정책적 방향을 개발하는 것이다. 주로 거시적 관점의 국가수준에서 어떻게 과학데이터 시각화를 체계적으로 지원할 것인가를 정책적인 시각에서 분석하였다. 구체적으로 기존 문헌의 조사를 통하여 주요 성공요인들을 파악하였으며, 집중적으로 투자해야 할 연구의 영역을 설정하였다. 다음으로 국가수준의 시각화 연구지원의 전략적인 방향과 전략적 과제를 제언한다. 제2장에서는 과학데이터 시각화 연구동향을 조사했으며, 제3장에서는 과학데이터 시각화를 위한 주요연구영역을 파악하고, 제4장에서는 연구개발지원 정책방향 및 과제목록을 개발하고, 마지막으로 제5장에

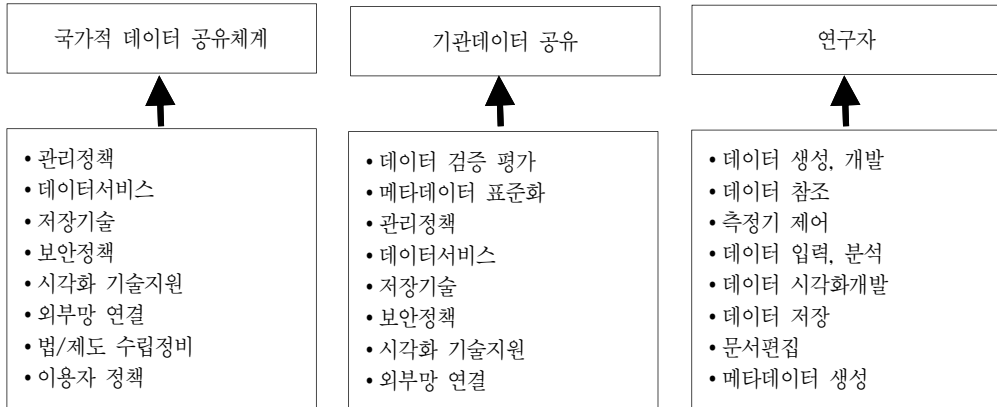
서는 결론을 서술한다.

2. 과학데이터 시각화 연구동향

2.1 e-science 기반의 연구지원시스템

연구자는 연구과정에서 연구목적에 위해 데이터를 다음과 같은 단계로 수집 관리한다고 볼 수 있다. 첫째, 자신의 컴퓨터에 연결된 계측장비와 특정 소프트웨어를 사용하여 수집(collection)한다. 둘째, 표준화되지 않은 임의의 절차에 따라 측정데이터를 저장하고 설명자료 추가하여 보존(curation)한다. 셋째, 데이터 수집이나 보관을 위해 표준화된 절차와 무관하게 분석용 소프트웨어에 맞게 가공하여 데이터를 분석(analysis)하게 되는데 주로 상용소프트웨어를 많이 활용한다. 넷째, 마지막 단계로서 해석을 용이하게 하기 위해 분석단계로부터 얻어진 결과를 시각화(visualization) 소프트웨어를 사용하여 도형, 그림, 영상, 그래픽 등의 시각정보를 변화하여 표시한다. 과학데이터를 국가적으로 관리하는 조직, 기관 공유의 데이터를 관리하는 조직, 데이터 생성자 이면서 사용자인 연구개발자 역할이 <그림 1>과 같이 정리된다(한국표준과학연구원 2010).

e-science 환경에서 과학데이터의 시각화는 <그림 1>에서와 같이 국가적인 수준에서 과학데이터를 공유하고 관리해야 한다. 과학데이터의 관리 및 제공은 이용자가 편리하고 유용하게



〈그림 1〉 e-science 환경

사용할 수 있도록 환경이 마련되어야 한다. 과학자가 원하는 과학데이터를 정확하고, 직관적으로 이해하기 위해서는 원시데이터 형태보다는 시각화된 환경에서 이용할 수 있도록 다양한 지원이 필요하다. 데이터를 표현하기 위해 사용되는 하드웨어 및 소프트웨어 환경과 시각화된 환경에서 데이터의 이해 방법, 활용방법 등의 정보가 제공되어야 한다. 과학데이터의 생애에서와 같이 데이터를 가공 분석하여 제공하는 시각화 기술은 매우 중요하며 이를 실현하기 위해서는 〈그림 1〉에서와 같이 국가적인 데이터 공유체계에서 하나의 구성요소로서 시각화 기술지원이 필요하다.

최근 과학데이터 시각화에 대한 연구가 활발히 이루어지고 있는 가운데 과학기술 출판물에서 시각화를 효과적으로 표현하기 위한 기준으로 정보전달을 위한 그래픽 단순화, 도표 등을 만들기 위해 사용하는 객체나 속성 형태에 대한 표현, 표현 목적에 따라 색상, 선, 연

결점에 대한 표현 등의 가이드라인을 제시하고 있다(Kelleher and Wagener 2011).

e-science 환경을 구현한 해외의 사례를 살펴보면 〈표 1〉과 같다. 이러한 사례의 공통점은 연구과정 시나리오를 바탕으로 개발되었으며, 서비스의 목적과 환경에 맞게 표준화된 기술을 사용하고 있다.

과학데이터를 이용해서 시각화를 구현할 경우 기초데이터가 바뀔으로써 자연스럽게 시각화된 내용도 최신성을 유지해야 함에도 불구하고 더 이상 데이터로서 관련되기가 어렵다. 이런 이유는 두 가지로 생각해 볼 수 있는데, 첫째 과학데이터의 수집의 문제이다. 과학데이터의 이용에 초점을 맞추어지기 보다는 데이터를 생성하는 장치의 구성이나 그 특정형태에 맞추어서 생성된다는 것이다. 둘째는 데이터의 규모로 새로운 데이터 수집 장치의 설계 능력이 지속적으로 향상될 뿐만 아니라 양적으로 풍부하고 질적으로 우수한 데이터의 요구

〈표 1〉 e-Science 환경 개발사례

| 프로젝트 | 개발주체 | 목적 | 서비스 | 관련기술/표준 | 개발개념 |
|-----------------------------|--------------|-------------------------------|---|---|------------------------|
| Harvard Dataverse & Network | Harvard 컨소시엄 | - 데이터공유 | - 데이터 소유권 - 데이터 인용성 - 데이터형식변환 - 접근성 제어 - 데이터 관리 | - DDI - Simple Dublin Core(DC) - CSDGM - UNF 인코딩 - DOA - Handle system | - Web - Open Source |
| eSciDoc | 독일 컨소시엄 | - Data publication - 데이터공유 | - 연구저작, 관리 - 데이터 공유 | - PID - OAI-PMH protocol - OAI-ORE protocol - shilbboleth LDAP - OpenSearch SRU/W - Fedora content model | - Web - Open Source |
| Dallas | Microsoft | - Data curation | - 데이터 저작 - windows Azure Plateform | - ontology Add-in 4 Office - authoring Add-in 4 word2007 - Cehm4Word: chem. drwaing in word | - Web - 상용 |

가 데이터 이용자인 과학기술자에 의해 요구되고 있다. 이러한 경향은 데이터를 기반으로 한 새로운 연구영역의 탐색보다는 전통적인 데이터 검색 효율성에 기초한 데이터베이스 기반 초점을 맞추어 왔기 때문이기도 하다(Fox and Hendler 2011).

e-science 환경을 실현하기 위한 중요한 과제는 객체 표현 기술에 있다. 연구결과는 데이터뿐만 아니라 그림, 텍스트, 실험장치의 규격, 시각화 소프트웨어의 명세서(Specification), 데이터 검색을 위한 메타데이터 등이 포함되어야 함으로 각각의 객체를 표현하기 위한 기술이 〈표 2〉와 같이 요구된다.

과학데이터를 국가적 수준에서 저장, 관리, 유통하기 위해서는 연구데이터만이 갖는 고유

한 메타정보를 구성하여 원시데이터 및 시각화 소프트웨어에 접근하기 전에 내용을 파악할 수 있도록 메타정보가 필요하며, 향후 과학데이터 시각화 분야에서 메타데이터 표준화가 이루어진다면 시스템 간의 정보자원의 상호운용성을 확보할 수 있게 될 것이다. 시각화 분야의 메타정보를 정의하기 위해서는 국제적 표준(<http://metadata-std.org/11179>)인 메타정보 등록 모델인 관리식별 기능, 데이터명 정의, 분류, 데이터 요소, 데이터 요소의 개념, 개념적 영역 등에 대한 검토가 필요하다. 미국 환경청은 환경관련 표준들을 기록하게 하여 기관과 사용자 간의 데이터 공유가 가능하게 하였으며, 이해가 쉽고 신뢰성이 높은 환경관련 자료를 제공하는 EDR¹⁾(Environmental Protection Agency

1) http://iaspub.epa.gov/sor_internet/registry/sysofreg/home/overview/home.do

〈표 2〉 e-Science 환경구축을 위한 플랫폼과 요소기술

| 개발항목 | 개발내용 | 비고 |
|-------------------------|--|------------------------------|
| e-Science 플랫폼 | 4단계 연구과정을 통합하고 데이터와 시각화 프로그램을 공유할 수 있도록 지원하는 시스템 플랫폼 | DOA, Fedora 참조 |
| 객체 표현기술 | e-Science 플랫폼을 구현하는데 필요한 요소기술. 디지털객체 식별자, 데이터 인용 식별자, 메타데이터 표현 인코딩 | DOI, UNF, PID 참조 |
| 시각화 프로그램 메타데이터 표현 | e-Science 환경에서 시각화 프로그램을 등록하고 인용하기 위해 메타데이터 표현 형식을 정의 | 시각화 메타데이터 항목 |
| 연구데이터 범용 데이터 포맷 및 뷰어 개발 | 연구 데이터의 특징은 이를 분석하는 소프트웨어 종류만큼이나 다양하고 이를 시각화하기 위해서는 각 소프트웨어의 데이터 포맷에 맞는 형식으로 입력을 수행해야 함. | DICOM, KML, CSV, Shapefile참조 |
| 시각화 프로그램 검증센터 | 연구데이터 시각화 프로그램의 개발과 공유를 지원하기 위한 조직구성과 기능 | 시각화 SW 검증센터 |

Metadata Registry)은 ISO/IEC11179를 기반으로 구축한 대표적인 사례이다. 과학데이터 및 시각화 프로그램의 메타데이터 항목을 표

준화하고 이를 전자적 방법으로 인코딩하기 위해 〈표 3〉과 같이 메타데이터 표준안을 제안한다.

〈표 3〉 과학데이터 시각화 메타데이터(안)

| 항목이름 | 설명 |
|------------------|-------------------------|
| 연구결과 인쇄물 제목 | 연구데이터 생성과 관련한 연구결과물의 제목 |
| 연구결과 인쇄물 고유번호 | 연구결과물의 등록번호 |
| 연구데이터 고유번호 | 연구데이터 번호 |
| 연구데이터 생성장치 명 | 연구데이터를 생성에 사용한 장치명 |
| 연구데이터 생성 장치의 제조사 | 제조사 이름 |
| 연구데이터 생성장치의 모델명 | 모델이름 |
| 연구데이터 생성장치의 특징 | 장치의 입출력 특징 |
| 연구데이터 분야 | 연구데이터의 분야 |
| 연구데이터 설명(초록) | 연구데이터의 설명 |
| 연구데이터의 내용 | 연구데이터의 구성내용 |
| 연구데이터의 형식 | 데이터 저장형식 |
| 시각화 프로그램 이름 | 프로그램 이름 |
| 시각화 프로그램 번호 | 프로그램 번호 |
| 프로그램 언어 및 라이브러리 | 프로그램 개발환경 |
| 시각화 프로그램의 요점 | 프로그램 개발 방법 |
| 시각화 프로그램의 버전 | 프로그램의 버전 |
| 시각화 프로그램의 목적 | 시각화 프로그램의 목적 |

2.2 과학데이터 시각화 프로세스

시각화 프로세스란 데이터가 시각화 과정을 거쳐 지식화되는 과정을 보여 주는 프로세스이다. 여기서 사용자는 적극적 참여자이며, 시스템 측면에서 보면 상호대화에 있어서 유연성이 공급되어야 한다. <그림 2>와 같이 시각화 프로세스는 비교적 간단하나 시각화 연구에 있어서 집중적으로 연구되어야 할 중요한 영역인 인식/인지, 조사/상호대화, 명세화/시각화 등에 관하여 보여준다(Wijk 2005).

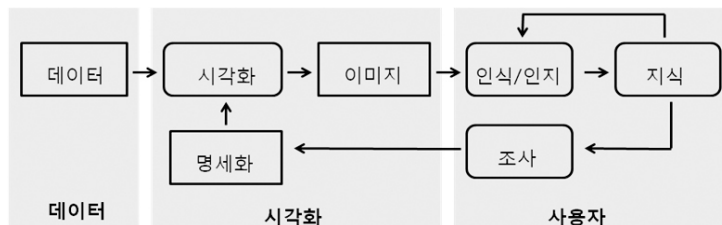
시각화에 사용되는 데이터세트는 하나의 비트 수준에서부터 시간에 따라 변하는 3D 데이터 등에 이르기까지 매우 다양하다. 시각화 명세화는 하드웨어, 알고리즘, 특정 파라메타 등을 포함한다. 사용자는 상호대화적인 방법으로 시각화의 명세화 값을 조정한다. 결과 이미지는 일반적으로는 이미지이나, 애니메이션, 소리 및 햅틱 피드백 등이 되기도 한다.

2.3 과학데이터 시각화 활성화 분야

시각화 전문가는 시각화 기술이 적용되는 다

양한 분야의 도메인 전문가와 밀접하게 협력함으로써 문제해결을 위해 현실적으로 필요한 툴과 기술 적용 및 연구개발을 할 수 있다(Johnson et al. 2006). 즉 시각화 학문분야가 발전하기 위해서는 다른 학문분야의 전문성과 결합하여야 한다. 또한 시각화 전문가들은 데이터를 분석을 위하여 시각화 기술을 통계적 처리방법, 데이터 마이닝, 이미지 프로세싱 등과 같은 다른 분석적 툴과 결합되기도 한다. Visual Analytics 가 한 예가 될 수 있다(Thomas, J. and K. Cook 2005).

시각화는 이미 의학, 과학, 공학, 비즈니스 분야에 지대한 영향을 끼쳤으며, 이들 분야는 물론 다른 모든 분야에 있어서 적용됨으로써 잠재적인 영향력은 매우 클 것으로 예상된다. 각 연구 영역에서 시각화는 매우 활발하게 연구되고 있으며, 특히 생명과학 분야는 시각화 연구가 상당한 수준으로 발전해 왔다. 예를 들면 바이오인포매틱스 시각화(Bioinformatics Visualization), 수술지원(Surgical Support), 예방과 정책(prevention and policy), 바이오 이미징(Biological Imaging), 개인맞춤처방(Personalized Medicine) 등의 분야이다.



<그림 2> 과학데이터 시각화 프로세스

2.4 해외 기술개발 현황

국가수준의 시각화 체제의 중요한 역할 중의 하나는 시각화 툴에 대한 지원과 교육이다. 특히 오픈소스 라이브러리를 구축하여 시각화 도구를 제공하는 것이 중요하다. 본 연구에서는 최근 주목받고 있는 시각화 분석 도구 및 서비스를 웹사이트를 중심으로 조사한다.

연구데이터 시각화시스템의 국외현황을 살펴보면, 분자생물학이나 바이오인포매틱스와 관련된 도구들 또는 이들 데이터베이스에 대한 웹 서비스가 증가하고 있는 추세이다. 대표적인 곳으로는 영국의 EMBL(European Molecular Biology Laboratory)-EBI(European Bioinformatics Institute), 미국의 NCBI(National Center for Biotechnology Information)²⁾, 일본의 DDBJ(DNA Data Bank of Japan)와 PDBJ(Protein Data Bank of Japan)³⁾가 있으며, 그 외에도 연구소나 프로젝트별로 다양한 웹 서비스를 제공하고 있다. 데이터 시각화 서비스의 경우 IBM의 Open Visualization Data Explore와 VIS OpenDX⁴⁾가 개발되어 활용 중에 있다.

심리·생리 데이터의 활용을 위한 측정과 분석에 관련하여서는 대표적으로 미국의 ADI Inc., BIOPAC 시스템, 유럽의 Brain Product, ANSLAB 등에서 각각의 프로그램을 개발하

여 시각화를 지원하고 있다. 기타 분야로 미국 LLNL(Lawrence Livermore National Laboratory)⁵⁾에서는 VisIt을 이용하여 주로 폭발 현상에 대한 시뮬레이션을 수행하고 있으며, NASA(National Aeronautics and Space Administration)에서는 시각화 시스템을 구축하여 우주선의 비행 이력에 대한 분석과 이의 시각화 서비스를 제공하고 있다.

시각화 센터의 운용과 관련해서는 미국의 경우 NVAC(National Visualization and Analysis Center)⁶⁾를 통하여 국가 차원의 시각화 분석 기술과 도구의 자원공급에 대한 전략적 지휘와 조정 역할을 하도록 하고 주요 대학에 지역 데이터 시각화 센터를 두어 이들을 연결하는 임무를 수행하고 있다.

3. 과학데이터 시각화 지원을 위한 주요영역 설정

본 연구에서는 우리나라 실정에 적합한 과학데이터 시각화를 지원하기 위해서 먼저 전략적으로 고려되어야 할 주요영역을 설정하였다. 이는 한국과학기술정보원, 한국과학표준연구원, 한국기초과학지원연구원 등의 연구원들과 “과학기술 연구자를 위한 연구데이터 시각

2) <http://www.ncbi.nlm.nih.gov/>

3) http://www.pdbj.org/pdbj_links_k.html

4) <http://www.opendx.org/>

5) <https://www.llnl.gov/>

6) <http://nvac.pnl.gov/centers.stm>

화 시스템 개발” 연구과제에 참여하고 있는 연구원들과의 5회의 브레인스토밍 과정을 통하여 도출되었다. 특히 주요 성공요인(CSF: Critical Success Factors)을 도출하기 위해서는 먼저 방대한 문헌 조사를 실시하였으며, 아래와 같이 가장 중요한 5가지의 CSF를 결정하고 관련된 문헌을 언급한다.

3.1 과학데이터 시각화 연구의 주요 성공요인

과학데이터 시각화와 관련된 문헌들을 조사하여 다음과 같은 5가지의 주요 성공요인을 파악하였다.

첫째, 과학데이터 리포지토리를 통한 데이터 및 시각화 툴의 공유이다. 과학의 기본적인 특성 중 하나는 같은 실험을 반복할 수 있다는 것이다. 시각화도 여러 연구자들에게 반복되어질 필요가 있는데 이를 위해서는 먼저 데이터가 공유되어야 한다(Fekete, J. 2004). 아울러 시각화 모델 및 시각화 알고리즘과 테크닉도 공유되는 것이 바람직하다. 특히 국가 R&D 예산 지원을 통한 연구에서 산출된 연구성과물을 공개하는 것은 공익의 목적을 위해 당연한 조치이다.

둘째, 국가수준의 연구지원 기관에서 리더십 발휘이다. 미국의 경우 2004년부터 NSF(National Science Foundation)와 NIH(National Institute of Health)에서 시각화 연구를 위한 Visualization Research Challenges Ex-

ecutive Committee를 발족하여 워크숍을 통하여 연구방향을 제시하고 로드맵 작성을 위한 연구를 진행하고 있다(Hansen & Johnson 2004). 우리나라에서도 시각화 관련기관에서 리더십을 발휘하여 시각화를 위한 연구지원체제, 주요연구주제, 로드맵 등을 작성하여 체계적으로 접근하여야 한다.

셋째, 인간의 인식/인지능력 반영하는 시각화 시스템 개발이다. 시각화 연구는 시각화의 효과성을 향상시키고 인간의 추론능력(reasoning capability)을 향상시키기 위하여 비주얼 디스플레이를 효과적으로 사용하는 것이 핵심이다(Ware 2000). 따라서 시각화가 언제, 왜, 어떻게 데이터 소스로부터 통찰력을 갖게 하는가에 대한 연구가 필요하다. 즉 인간의 인식, 인지, 시각화 경험 등에 대한 특성 견해 및 제한성을 파악하는 것이 시각화와 비주얼 커뮤니케이션에 있어서의 발전을 촉진할 것이다.

넷째, 다양한 데이터 표현을 위한 신기술 개발이다. 현재의 시각화 기술은 매우 강력하지만 아직도 완벽하지는 않다. 특히 데이터 소스가 계속적으로 다양한 형태로 확장되기 때문에 이들을 동기화(synthesis)하고 표현(representation)하기에 어려움이 많다. 따라서 복잡하고, 다변량적(multivariate)이고, 이질적(heterogeneous)이며 다중스케일(multiscale)이거나, 동적인 데이터를 위한 표현방법에 대한 신기술의 발전이 계속적으로 필요하다(Hey et al. 2009).

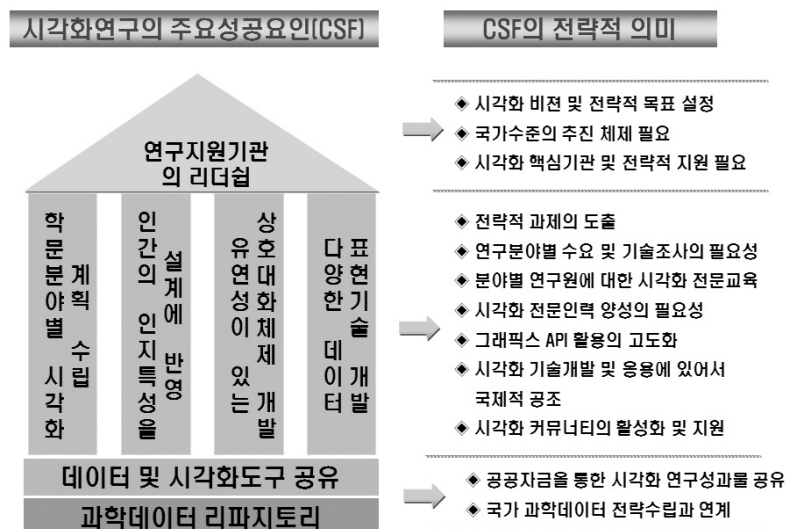
다섯째, 학문분야별 시각화 계획수립이다. 같은 학문분야의 시각화 기술에 대한 요구사항은 거의 동일하다. 즉 과학데이터를 공유함과 같이 시각화와 관련된 일련의 기술 및 방법론도 공유하는 것이 바람직하다(Johnson et al. 2007). 따라서 각 학문분야에서는 그 분야에 맞는 시각화 기술개발 및 적용에 대한 체계적인 계획을 수립하여 자원을 효율적으로 배분하는 것이 중요한데, 구체적으로 현황실태조사, 요구사항 및 수요조사, 기술조사, 그리고 시각화 관련 프로젝트 등에 대한 계획을 수립하여 국가수준의 연구지원 기관과 협력하여 진행하는 것이 필요하다.

이상에서 언급한 주요 성공요인과 이들의 전략적 의미를 정리하면 <그림 3>과 같다.

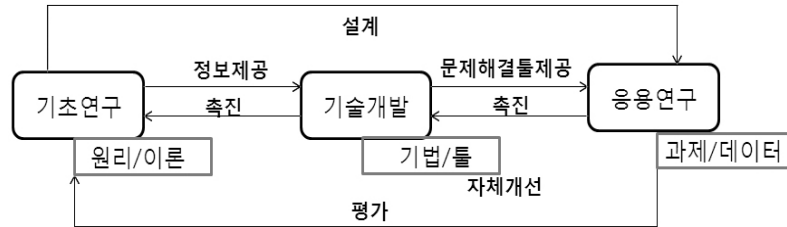
3.2 주요 연구영역의 설정

일반적으로 과학기술에 있어서 특정분야의 연구의 영역은 <그림 4>와 같이 크게 세 가지 카테고리로 나눌 수 있다. 즉 기초(basic)연구, 기술개발(transitional)연구, 응용(applied)연구로 구분된다. 기초연구는 특별한 응용 또는 사업을 직접적으로 목표로 하지 않고 새로운 지식을 획득하기 위하여 최초로 행해지는 이론적 또는 실험적 연구를 말한다. 기술개발연구는 기초연구의 결과물을 활용하여 장치 현실문제 해결을 위한 기술을 개발하는 것이다. 응용연구는 기초연구를 통해서 개발된 이론과 개발된 기술을 이용하여 실용적인 목표 하에 새로운 과학적 지식 혹은 문제를 해결하기 위한 연구를 말한다.

시각화에 있어서 기초연구 영역에는 시각화



<그림 3> 주요성공요인과 전략적 의미



〈그림 4〉 과학기술 연구 영역과 관계성

와 관련된 기본적인 컴퓨터이론과 함께 정신 물리학(psychophysics), 지각(perception), 그리고 비주얼 코딩 등을 포함한다. 반면에 응용 연구에는 유체역학, 메디컬 이미지, 정보통신 네트워크, 소프트웨어공학 등과 같은 특정 도메인의 과업들과 이들과 관련한 데이터들을 포함한다. 그러나 시각화 측면에서 가장 활발한 연구는 이들 두 카테고리 사이에 있는 기술 개발 영역에서 수행된다. 이는 시각화 분야는 신이론을 개발하는 것보다 기술개발에 치중하기 때문이다.

이상의 시각화 연구영역은 각각 카테고리들 간에 복잡한 사이클 관계를 보인다. 〈그림 4〉에서와 같이 기술개발 연구영역에서의 자체개선(refine)은 해당 영역 안에서의 피드백에 따른 개선연구 활동이다. 기타 모든 화살표로 이루어지는 사이클은 각 영역들 사이의 관계를 보여준다. 예를 들면 기초연구의 결과(성과)는 기술개발연구의 기반이 되는 이론들이 되며, 반대로 기술개발연구는 기초연구를 유도(촉진)하는 요인이 되기도 한다. 기술개발연구의 성과는 응용문제들을 해결하는 기술들이 될 것이며, 이것은 새로운 기술 혹은 더 나은

기술의 개발을 유도할 것이다.

응용연구는 자체의 사이클들이 있는데, 한 쪽 방향으로서는 기초연구와 기술개발연구에서 응용문제 해결을 위한 지식을 제공할 것이다. 반대방향으로는 응용문제에 대한 솔루션을 평가하면서 왜 특정한 접근방법이 의도하는 문제를 푸는지 혹은 풀지 못하는지에 대한 원인을 분석함으로써 기술개발 단계의 기술의 개선과 기초연구단계의 이론을 다시 검토하도록 유도한다.

이러한 관계를 기반한 사이클들을 통하여 시각화 분야의 연구가 탄력을 유지할 수 있도록 기획하고 연구지원이 이루어지도록 하는 것이 매우 중요하다. 그러나 기술개발(기술화, 산업화)연구가 시각화 연구에 있어서 핵심연구영역 일지라도 아직 많은 투자가 이루어지고 있지 않는 실정이다. 결론적으로 시각화 연구자 혹은 그룹의 이상적인 연구 프로그램은 기초연구, 기술개발연구, 응용연구를 타겟팅 하면서 〈그림 4〉의 모든 사이클을 포함해야 한다.

4. 연구개발지원 정책방향 및 과제

국가적 수준에서 과학데이터 시각화의 연구 개발 지원 및 정책 방향에 대한 연구는 시각화 응용소프트웨어 개발, 시스템 평가, 질량분석 연구자 등 관련 분야 전문가의 자문과 심층 토론으로 도출된 것이다. 객관적인 절차와 방법론을 도입하지 않아 다소 주관적인 면이 있어 연구의 한계라고 볼 수 있다. 따라서 이 부분에 대해서는 추후 객관적인 연구방법에 기초한 심도 있는 연구가 되어야 한다. 그러나 본 연구에서는 국내에서 아직까지 시도하지 않은 과학데이터 시각화 관련 정보 및 기술을 연구 개발자에게 제공하기 위한 방안을 제안하는 측면에서 의미를 둔다.

4.1 수요지향적 시각화 R&D 전략 및 정책

4.1.1 활성화된 연구영역의 과학데이터 확보 및 리포지터리

현대 과학의 패러다임은 데이터베이스를 기반으로 새로운 사실을 발견하는 방향으로 전개된다. 따라서 각 학문분야에서는 다양한 데이터의 확보와 아울러 적용 가능한 데이터 분석기법들을 도입하고 있다. 이때 시각화는 대용량의 데이터베이스를 분석하는 데 있어서 가장 편리하며 필수적인 방법이다.

시각화를 통해서 새로운 지식의 발견을 위해서는 먼저 과학데이터의 생산, 큐레이션 및

유지관리가 무엇보다 필요하다. 따라서 주요학문분야의 데이터 센터에서는 시각화를 위해서 데이터를 확보해야 할 뿐 아니라 오픈소스화된 시각화 툴을 저장하여 제공하는 것이 바람직하다. 아울러 각 학문분야에서는 필요한 시각화 툴을 개발하거나 구매하여 데이터센터에서 데이터와 같이 제공해 주는 것도 중요하다.

4.1.2 학문분야 연구문제에서 접근

과학적 연구문제 해결에 있어서 시각화는 연구의 전 과정에 적용된다. 구체적으로 실험과정의 평가, 실험결과 해석, 실험결과 신뢰성분석 등에서 데이터의 의미를 해석하기 위하여 다른 데이터 분석 툴과 같이 사용된다. 실험과정의 평가에서는 주로 각 단계의 적합성 평가에 데이터 시각화 기술이 사용될 수 있다. 실험결과 해석에서는 시각화를 통해서 실험결과 질적 향상을 파악할 수 있으며, 새로운 지식을 추출하는 데 기여한다. 신뢰성 분석에서 사용되는 통계적 기법의 적용 결과 분석에 시각화 기술들이 적용된다.

연구분야별로 전형적인 실험의 방법과 절차에 맞는 시각화 기술들을 분석 정리하는 것이 필요하다. 해당 분야별로 다양한 시각화 기법들이 적용될 것이므로 각 분야의 전문가와 함께 이러한 기술들을 파악하고 확보할 필요가 있으며, 개발할 필요가 있는 시각화 기술을 확보하기 위한 연차적 계획이 수립되어야 한다.

4.1.3 협업 연구에 있어서 시각화를 위한 인프라 구축

하드웨어 특히 CPU와 그래픽 디스플레이 기술의 발달 - 특히 Volume rendering을 위한 하드웨어의 발달 - 등으로 인해 하드웨어적으로는 국가수준에서 연구를 위한 자원투입의 필요성은 상대적으로 매우 작다. 그러나 대용량 데이터 세트를 로컬에서 시각화를 하기 위해서는 고속의 네트워크가 있어야 하며, 아울러 원거리에서 서버에 있는 대용량의 데이터를 시각화할 수 있는 네트워크 인식(network-aware) 시각화 알고리즘의 개발이 필요하다. 이러한 기술은 원격에서의 시각화와 협업연구가 가능하게 한다.

4.1.4 시각화 소프트웨어 리포지터리 및 유통체계의 구축

시각화 시장이 성장하면서 상업화 및 공개용 시각화 툴과 소프트웨어가 많이 출시되고 있다. 특히 공개용 오픈소스 모델은 학계 및 산업계의 연구자와 최종사용자들이 유용하게 사용할 수 있다. 이와 관련된 정보와 툴을 제공하는 것은 매우 중요하다. 공개된 시각화 툴과 아울러 개인 혹은 특정 연구팀에서 연구수행과정에서 개발한 시각화 툴 및 분석방법을 공개하여 리포지터리에 저장할 수 있는 체계가 필요하다. 구체적으로 시각화 툴 및 데이터의 메타데이터, 마크업 언어, 검증절차, 인준절차, 등록을 위한 절차와 보상체계 등에 대한 연구가 필요하다. 국가적으로 과학데이터의 유통과

아울러 해당데이터를 분석할 수 있는 시각화 툴을 소개하고 이를 유통시킬 수 있는 체계가 필요하다.

4.1.5 그래픽스 API 응용프로그램 개발에 중점

〈그림 4〉에서 언급된 것처럼 시각화 연구에 있어서 가장 핵심적 연구는 기술개발(transitional) 분야이다. 기술개발은 크게 두 가지로 구분되는데 첫째는 그래픽스 API를 개발하는 것이고, 둘째는 그래픽스 API를 이용하여 과학기술 분야에 적합한 시각화 응용 프로그램을 개발하는 것이다. 응용프로그램과 그래픽 하드웨어의 사이에 존재하는 API(고수준, 저수준)는 이미 국제적으로 표준화 되어 있으며 많은 기술들이 개발되었다. 우리나라의 수준에서 보면, 그래픽스 API 개발을 위한 연구인력과 자원은 절대적으로 부족하나, 많은 API 들이 국제적으로 표준화 및 상용화된 툴로서 제공되고 있으므로 이에 대한 연구자원의 투입 필요성은 상대적으로 적다. 그러나 이러한 API 기술을 이용하여 연구자들이 자신의 연구를 진행하면서 필요로 하는 응용프로그램의 개발에 대한 투자가 필요하다. 일반적으로 도메인 연구자들은 이러한 기술의 활용과 프로그래밍에 제한적이므로 연구자들을 지원하면서 같이 응용프로그램 개발을 지원하는 체계가 있어야 한다.

한 예로 1992년 실리콘 그래픽사에서 만든 3차원 그래픽 표준 API 규격인 OpenGL(Open Graphics Library)은 프로그램 언어, 플랫폼

간의 상호 응용 프로그램을 지원한다. 250여개의 함수호출 기능으로 단순한 기하도형과 복잡한 3차원 화면을 구성할 수 있다. 이 표준 API는 가상현실, CAD, 정보시각화, 비행시뮬레이션 등에 활용되고 있으며 게임 산업에도 많이 활용되고 있다(위키백과 2011).

이와 같이 시각화 관련 표준 API는 이미 많이 개발되어 활용되고 있지만 국가적 수준에서 시각화 응용프로그램을 지원하기 위해서는 공개된 정보원을 체계적으로 수집할 필요가 있다. 사용법, 사용규약, 활용 방안 등에 대한 내용을 정리한 리포지토리를 구현하여 제공하는 것도 의미 있는 서비스라고 할 수 있겠다.

4.2 전략적 과제

4.2.1 시각화 거버넌스 국가모델의 개발

장기적으로 시각화에 대한 국가 전략은 국가 과학데이터 관리와 같은 맥락에서 고려되어야 한다. 따라서 시각화 비전과 전략적 목표 설정, 사업기획 및 실행계획 수립은 현재 추진 중인 “국가 과학데이터관리를 위한 전략계획”의 결과를 참조하여야 한다. 이를 통하여 시각화의 TO-BE 모델을 개발하여야 한다.

국가적으로 과학데이터를 이용한 시각화의 혜택을 최대화하기 위해서는 핵심기관과 관련 기관 및 연구원들의 조정과 협력이 필수적인데 이들의 역할을 체계화하는 모델의 개발이 필요하다. 타당성 조사 수준인 본 연구의 결과

물을 기반으로 구체적인 모델의 개발이 필요하다.

국가수준에서 시각화 연구는 국가적인 연구 협력체제 하에서 이루어져야 한다. 즉 기존의 인프라와 각 조직의 특성 및 장점을 반영하고, 정보서비스는 기존의 정보서비스 체제와 연계하여 구축하여야 한다.

4.2.2 시각화 센터의 설립 및 운영

국가수준에서의 과학데이터 시각화 관리계획을 수립하고 실행하며 시각화 도구들을 관리하고 장기보전하며 서비스하는 전문단체가 설립되어야 하는데, 이를 위해서 먼저 법과 제도를 정비할 필요가 있다. 구체적으로 최고의사결정을 하는 상위수준의 위원회가 필요하며 이를 지원하는 국가 시각화 센터(National Visualization Center)가 설립되어야 한다.

또한 기능적으로는 시각화 리포지토리를 개발해서 운영하고, 시각화 기술의 유통과 정보 서비스를 제공하며, 큐레이션과 장기보전을 위한 인프라를 확보하여야 한다.

무엇보다 중요한 것은 각 연구자들에게 시각화 기술 카운슬링을 할 수 있는 인력을 확보하여야 한다. 이들은 도메인에 전문성이 있는 연구자와 함께 해당분야의 문제해결을 위한 프로그래밍을 하는 것이다. 추가적으로 국제적인 시각화 기술동향 및 벤치마킹을 실시하고 국제단체와 교류 및 협업하는 체제를 구축하는 기능도 포함된다.

4.2.3 분야별 시각화 기술 수요조사 및 기술 동향분석

시각화 연구 자원의 선택과 집중을 위해서는 각 분야별로 관리되거나 개발되어야 할 기술 혹은 시각화 툴에 대한 조사가 선행되어야 한다. 각 학문분야는 비슷한 연구프로세스와 해당 단계별로 요구되는 시각화 기술들이 있다.

수요조사를 통해서 기존의 일반형(general-purpose) 시각화 툴 및 맞춤형(customized) 툴, 오픈소스 프로그램 등을 조사하고, 분야별로 새롭게 개발되어야 할 시각화 기술들을 파악한다. 아울러 각 국제적으로 각 분야에서 어떠한 시각화 기술들이 적용되고 있는가를 파악하는 것도 개발계획 및 확보계획을 수립하는 데 있어서 중요한 요인이다. 시각화 툴을 많이 사용하는 학문영역을 계속적으로 조사하여 최신기술의 획득 및 개발대상을 결정한다. 특히 의료분야를 포함한 생명과학 도메인과 물리학과 지구과학 등의 세부분야 등도 시각화 기술의 개발이 필요한 중요한 학문영역이다.

4.2.4 선도 프로젝트의 실시

시각화 실행계획을 구체화하는 선도과제(pathfinder project)를 실시한다. 즉 가시적 결과가 예상되는 소규모의 프로젝트를 실시하여 일정 수준 이상의 성과를 보임으로써 첫째, 계속 연구의 기반을 구축하고, 둘째, 시각화 기술력을 확보하며, 셋째, 시각화 기술개발 분야

의 인식을 고조시킬 필요가 있다.

국가차원에서 시각화 연구는 중장기적 계획 하에서 점진적으로 추진되어야 하는데 pathfinder 접근방법은 위험을 최소화 시키며, 기술력을 확보하기에 적합한 추진 방법이다. 대신 pathfinder 프로젝트는 조심스럽게 선택되어야 하는데, 기존에 과학데이터가 확보되어 있어야 하며 데이터 관리 및 시각화 툴의 개발에 이미 학습이 되어 있는 조직이 바람직하다. 따라서 이런 차원에서 보면 이미 인프라가 구축되어 있는 KISTI의 사실정보데이터베이스의 한국인의 인체모형데이터와 한국표준과학연구원의 국가참조표준센터에서 pathfinder 프로젝트를 실시하는 것이 바람직하다.

4.2.5 시각화 전문가 양성 프로그램의 개발

현실적으로 연구과정에서 시각화는 연구자의 몫이다. 그러나 연구자가 시각화 툴을 사용하고 추가적인 프로그래밍을 실시하는 어려움이 많을뿐더러, 시각화 과정을 통하여 연구자가 본연의 임무에 투자하고 소요되어야 할 자원을 효율적으로 사용하지 못하는 경우가 발생된다. 따라서 이러한 문제를 해결하기 위해서 두 가지 관점에서의 노력이 필요하다. 첫째는 특정 분야의 연구자에게 빈번하게 사용하는 시각화 툴을 교육하는 것이다. 즉 연구자들이 시각화 툴을 사용하여 그들의 문제에 접근하게 하는 것이며, 둘째는 주요 시각화 툴에 대하여 사용은 물론 추가적인 프로그래밍도 가능한 전문가를 양성하는 것

〈표 4〉 전략적 과제

| 분야 | 세부과제 |
|--------------------------|---------------------------|
| 시각화 거버넌스 국가모델의 개발 | 비전 및 전략적 목표의 설정 |
| | 모델의 개발 |
| | 협력체제 구축 |
| 시각화 센터의 설립 및 운영 | 법/제도 정비 |
| | 국가 과학데이터 시각화 위원회 및 센터의 설립 |
| | 시각화 리포지터리의 구축 시각화 카운슬링 |
| 분야별 시각화 기술 수요조사 및 기술동향분석 | 시각화 수요조사 |
| | 시각화 기술조사 |
| 선도 프로젝트의 실시 | 선도과제의 계획수립 |
| | 선도과제의 선택 |
| 시각화 전문가 양성 프로그램의 개발 | 도메인 전문가 대상 시각화 톨 교육 |
| | 특정 시각화 톨 전문가 교육 |

이다. 해당 시각화 톨의 전문가는 도메인 연구자와 협력하여 연구에 있어서 시각화를 실시하는 것이다.

본 소절의 내용인 전략적 과제를 요약하면 〈표 4〉와 같다.

5. 결론

지금까지 과학데이터에 대한 시각화는 개인 연구자 수준 혹은 학문분야 내에서 필요에 따라 실시되었다. 즉 시각화를 위한 투자는 제한적일 수밖에 없으며 체계적인 기술개발과 아울러 기술의 공유가 매우 힘들었다. 따라서 최근에 국가적 차원에서 시각화 기술을 개발하고 이를 유통 및 공유함으로써 각 학문 분야의

연구원들에게 시각화를 체계적으로 지원할 수 있음을 인식하였다.

본 연구는 이를 위하여 국가적 수준에서 전략적 방향을 제시한 것이다. 본 연구의 기대효과는 다음과 같다. 먼저 거시적 관점에서 보면, 첫째, 과학데이터 관리의 중요성이 부각되고 있는 현재 시점에서 과학데이터 분석을 위한 체계적인 지원의 토대를 제공한다는 것이다. 둘째, 연구자의 몫으로만 되어 있었던 과학데이터 시각화를 체계적인 지원의 필요한 국가적 과제로 관점을 바꿀 수 있는 중요한 계기가 되는 논문이다. 셋째, 시각화 분야에 있어서 미국 등 선진국의 연구 및 지원수준에 근접할 수 있는 정책적 방향을 제시했다는 것도 중요한 의미가 있으며, 이를 이용하여 정책 및 시각화를 위한 프로그램의 개발이 용이하

게 될 것이다. 다음으로 미시적 관점에서 보면, 첫째, 주요 성공요인과 전략적 의미의 도출사항은 시각화를 위한 세부지원사항을 구체적으로 파악할 수 있는 기준 및 개념적 틀이 될 수 있다. 둘째, 주요연구영역은 제한된 연구자원을 효율적으로 배정할 수 있는 기준이 될 수 있다. 셋째, 전략적 과제와 세부과제는 추후 국가적 시각화 지원사업의 구체적인 내용들이 될 수 있다. 마지막으로 본 연구는 국내에서 처음으로 국가수준의 시각화 전략을 논의한 탐색적 연구이므로 지속적인 연구의 단초를 제공한다는 측면에서 의미를 찾아볼 수 있다.

그러나 시각화 연구분야는 학문적인 분야로 완전한 체계를 이루지 못하고 있어 본 연구에서 주장하는 바에 대한 객관적인 자료 및 증거가 많이 부족하다. 특히 CSF 및 전략과제들의 도출은 브레인스토밍을 실시하였으나 방법론의 적용도 미흡한 것이 본 연구의 한계이다.

참고문헌

- 한국표준과학연구원. 2010. 『과학기술연구자를 위한 연구데이터 시각화시스템 개발』. [대전]: 동연구원.
- Christa Kelleher and Thorsten Wagener. 2011. "Ten guidelines for effective data visualization in scientific publications." *Environmental Modelling & Software*, 26(6): 822-827.
- Fekete, J., The Info Vis Toolkit, Proc. 2004. IEEE Symposium on Information Visualization (InfoVis). [cited 2010. 11. 22]. <<http://www.cs.umd.edu/hcil/iv04contest/>>.
- Hansen, C. and C. R. Johnson. 2004. *The Visualization Handbook*. San Diego: Academic Press.
- Hey, T., S. Tansley, and K. Tolle. 2009. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research.
- Johnson, C., R. Moorhea, T. Munzner, H. Pfister, P. Rheingans, and T. Yoo. 2006. *NIH-NSF Visualization Research Challenges Report*. Los Alamitos, CA: IEEE Computing Society.
- Johnson, C., R. Ross, S. Ahern, J. Ahrens, W. Bethel, K. L. Ma, M. Papka, J. van Rosendale, H. W. Shen, and J. Thomas. 2007. "Visualization and Knowledge Discovery: Report from the OE/ASCR Workshop on Visual Analysis and Data Exploration at Extreme Scale." [cited 2010. 11. 15]. <<http://www.sci.utah.edu/vaw2007/DOE-Visualization-Report-2007.pdf>>.
- Jarke J., Van Wijk. 2005. "The Value of

- Visualization.” In *Proc. IEEE Visua-*
lizatio, 2005: 567-574.
- Peter Fox and James Hendler. 2011. “Changing
the Equation on Scientific data Vi-
sualization.” *Science*, 331(6018): 705-
708.
- Thomas, J. and K. Cook, editors. 2005.
*Illuminating the Path: The Research
and Development Agenda for Visual
Analytics*. Los Alamitos, CA: IEEE
Computing Society.
- Ware, C. 2000. *Information Visualization:
Perception for Design*. San Francisco,
CA: Morgan Kaufmann.