

지구 환경 자료 분석을 위한 GeoGrid시스템 설계

Design of GeoGrid System for the Data Analysis of Global Environment

김태민*, 이명규, 최재영, 양영규

Tae-min Kim*, Myung-kyu Lee, Jae-young Choi, Young-kyu Yang

경원대학교 전자계산학과

scc0309@paran.com, bobbin@hanmail.net, {jychoi, ykyang}@kyungwon.ac.kr

요약

전 지구적 기온상승으로 인하여 해수면 상승 및 자연생태계가 파괴되고 있다. 이에 Geo 관련 연구자들은 지구 온난화, 이산화탄소 배출량, 미세먼지 확산 등의 방대한 지구 환경정보를 공유 및 교환하고 활발한 지식 교류를 하기 위한 포털 환경이 요구되고 있다. 국외 여러 환경단체들은 이러한 포털 환경구축을 통해 지구 환경정보를 연구하고 있다. 한국에서도 지구온난화 영향분석, 미세먼지 확산분석, 이산화탄소 배출량 분석, 지상 식생변화 모니터링 등의 분야에서 위성 관측 영상 및 지구환경자료를 이용한 서비스가 더욱 절실하다. e-Science 환경 구축의 대표 응용 분야 중의 하나로 발전, 확산시킬 필요성 때문에 정확하고 방대한 지식정보의 신속한 처리 및 서비스를 할 수 있는 GeoGrid 기반 기술 개발 방법과 환경 구축 방법에 대해 연구하였다.

핵심 용어 : GeoGrid, 지구환경, 온난화

1. 서론

최근 지구 온난화에 따른 해수면의 상승, 식량과 수자원 공급 부족, 인류의 건강 침해, 자연 생태계 파괴, 에너지 고갈 등 지구 생태계의 불안정한 요소를 발생시키고 있어 세계 각국의 관심 및 이에 대한 적응과 대응이 시급한 시점이다.

이에 각국은 온실가스 배출량 관리시스템을 구축하고 기업의 자발적 온실가스 감축 기반을 확충하는 등 다양한 방법들을 시도하고 있다. 하지만 아직도 지구 온난화 기후변화 감시와 지구온실가스 측정 및 분석을 위한 체계와 전략이 미비한 실정이다.

따라서 현시점에서는 개별적으로 관리·제공되고 있는 국내외 지구관측 분야별 자료를 통합하여 관련 연구 분야에서의 유기적으로 활용이 가능한 체계적인 환경 구축이 무엇보다도 시급하다. 즉, 관련 기관간의 긴밀한 협업 연구 환경 조성, 자원

과 자료 공유를 위한 시스템 개발 등의 전략적 계획과 함께 국가차원의 지속적인 지원이 필요하다. 이러한 지구환경 자료 등은 개별 대학이나 연구 기관이 독자적으로 보유, 관리하기가 힘들며 GRID 기반 연구 환경 구축 및 서비스 제공 기술개발을 통한 자료의 체계적 정리 및 공유가 필요하다.

이에 본 논문은 지구 온난화 분석, 미세먼지/황사 분석, 지질 정보, 지하자원 분석 등과 같은 분야에서 지구환경자료를 이용한 서비스 제공을 위해 국제적으로 운용하고 있는 지구환경 관측 자료에 대한 공유 플랫폼을 국내 환경에 맞게 적용하고자 한다.

또한, 이를 바탕으로 하여 국내·외의 지상 관측 등을 통한 지구환경 관측자료, 지리 정보 등의 데이터를 그리드를 통하여 효율적으로 처리하고 이를 효과적으로 활용하여 서비스 할 수 있는 GeoGrid 프레임 워크를 설계하고 개발하는데 그 목적

이 있다.

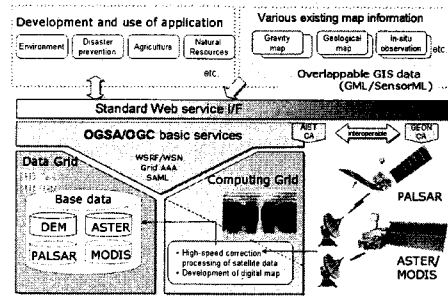
장기적인 차원에서 본 논문은 GeoGrid를 최대한 활용할 수 있는 잠재 수요 기관을 파악하여 사용자 그룹을 정의한 뒤 응용 연구 분야의 수요에 따른 정보·자료를 지속적으로 확보함으로써 활발한 커뮤니티를 형성할 수 있는 포털 환경을 제공하여 협업 환경에서의 연구 효율성 증대와 관련 연구 분야의 시너지 효과를 높일 수 있는 새로운 지구환경서비스 창출에 기여하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 지구 환경 데이터를 분석하는 국내외 기술 현황과 국내외 문제점을 제시하고 3장에서는 국외의 환경 데이터 분석 방법론에 근거하여 GeoGrid시스템의 분석과 설계를 한다. 4장에서는 기대효과와 결론을 기술한다.

2. 국내외 기술 현황

유럽 공동체가 지원하는 EGEE(Enabling Grid for E-science)는 대규모의 실용가능한 수준의 e-Science를 위한 그리드 인프라 스트럭처를 구축하여 새로운 산업계와 과학계의 자원과 사용자를 유치하고 있고, 영국은 국가 e-Science 센터와 전국에 10개의 지역 e-Science 센터를 구축하여 이들을 중심으로 산학연이 공동연구를 수행하는 체제를 구축하고, 이를 통해 많은 대형 협동 연구들을 수행 중에 있다.

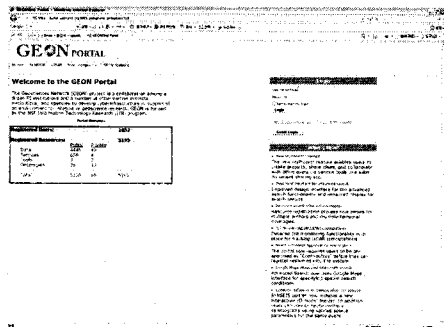
일본 AIST(National Institute of Advanced Industrial Science and Technology)를 중심으로 위성(ASTER)이 관측한 데이터들과 지질과학 정보, 환경 기술정보 등을 분석하는 GeoGrid(Global Earth Observation-Grid) 시스템을 구축하고 있다. <그림 1>은 Geo-Grid의 전체적인 개요를 나타낸다.



<그림 1> Geo Grid 개요

GEO(Global Earth Observation)-Grid는 지구환경 조사의 전체적인 데이터에 기반을 둔 GEO 과학의 기반환경을 구축한 것이다. Geo-Grid는 가상적으로 데이터들을 통합하고 사용자 하여금 Grid와 웹서비스 기술을 사용하여 통합된 데이터들에 대한 접근을 용이하게 해 주는 시스템이다. 이 시스템은 환경 보존, 자원 탐사, 자연 재해 예방 등의 범세계적인 사회 문제를 해결하는데 기여하는데 목적을 두고 있다.

미국의 GEON(Geosciences Network) 및 NEON(National Ecological Observatory Network)은 범지구적 기후 변화, 지질 변화 등이 생태계에 주는 영향을 파악하기 위하여 분산된 센서 네트워크와 실험, 표준화된 규약과 공개된 데이터 정책을 사용하여 연구 중에 있다. <그림 2>는 GEON 포털의 모습을 보여준다.



<그림 2> GEON portal

GEON 분산 시스템은 SOA(service-oriented architecture)에 기반을 둔 시스템

이다. 진보된 정보 기술은 지능화된 정보 검색, 데이터 통합, 그리고 비주얼화를 위해 개발되어 왔다. 환경은 고성능 컴퓨팅 플랫폼을 제공하는데, 이 플랫폼은 데이터를 강력하게 분석할 뿐만 아니라 계산중심의 모델 실행을 가능하게 한다. GEON 포털은 사용자들에게 편리한 웹 기반의 인터페이스를 제공하여 이런 기능들을 쉽게 이용하도록 해 준다.

국내는 2001년부터 정보통신부와 한국과학기술정보연구원(KISTI)이 주축이 되어 국가그리드 기술개발과제를 수행하였고, 현재 그리드 포럼을 운영함으로써 e-Science 핵심 기술 확보가 용이한 상태이다. 세계최고 수준의 IT인력과 기술력을 통하여 IT 기반 연구 생산성을 올릴 수 있는 e-Science 환경을 짧은 시간 내에 갖출 수 있는 잠재 능력 보유하고 있다.

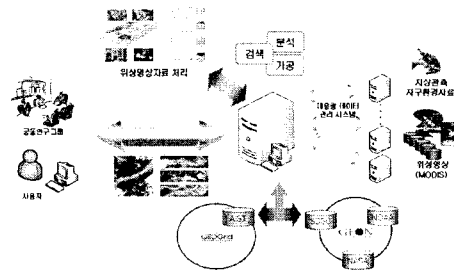
우리나라는 선진국과 비교하여 지구환경분야에서의 그리드 기술 활용 미미하여 지구환경분야의 협업 환경 구축의 부재로 인한 중복연구 및 투자가 요구된다. 또한, 일반적으로 지구환경 연구자의 e-Science에 대한 이해 및 인식 부족하여 지구환경분야에 있어서 최신 IT 기술 활용 부족하고 e-Science 분야의 전문 인력, 자원 및 예산지원의 부족하다.

e-Science 구축에 대한 수요조사 결과에 따르면, 우리나라 과학자의 90% 이상이 협업연구 및 연구자원에 대한 필요성을 느끼고 있으며, 이를 위한 IT 기술개발을 필요로 하고 있다. 지구환경 과학자들을 포함하여 e-Science 의존도가 높아질 것으로 전망하고 국가 대형 자료 공유에 대한 유기적인 네트워크를 형성하는 통합 e-Science 시스템 구축에 기여할 것으로 기대된다.

3. GeoGrid 시스템 분석 및 설계

GeoGrid 시스템은 <그림 3>과 같이 지구환경 관측자료 및 지구관측 위성영상을 관리 및 제공하며, 이를 이용한 서비스를 통하여 자료의 공유·관리 이외에도 이를

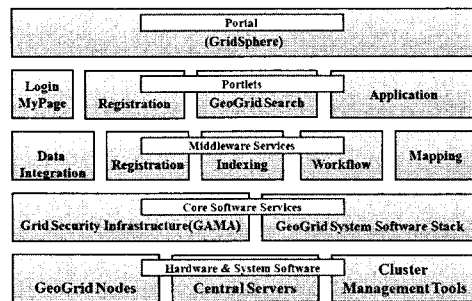
분석하고 의견을 교환하는 커뮤니티 장을 마련하는 Cyber-Infra 환경을 제공하고자 한다.



<그림 3> 개발 시스템 개요

3.1 GeoGrid 시스템 구조

<그림 4>는 GeoGrid 전체 시스템 구조를 보여준다. GeoGrid 시스템은 계층적 구조를 가지게 되며, 5단계의 계층으로 이루어져 있다.



<그림 4> GeoGrid 전체 시스템 구조

GeoGrid 구조는 하나의 GeoGrid 포털과 다수의 분산된 GeoGrid 사이트들로 이루어져 있다. 포털은 이 원격으로 이루어진 분산 환경에 접근할 수 있도록 하는 역할을 한다. 각각의 계층에 대한 설명은 다음과 같다.

가. Portal

GeoGrid 포털은 GridSphere Framework를 이용하여 개발된다. GridSphere는 포털 구조를 구현하고 포털 API와 호환된다. 많은 사이버 인프라 프로젝트가 포

탈을 구현하는데 GridSphere를 채택하고 있다.

나. Portlets

GeoGrid 응용 프로그램들은 사용자로 하여금 포탈에 있는 포틀릿으로 하여금 응용 프로그램을 사용할 수 있게 한다. 포틀릿은 서버쪽에서 담당하는 기능들에 대한 유저 인터페이스를 제공한다. 크게 두 가지로 나누어 보면 핵심 포틀릿과 응용 포틀릿으로 나눌 수 있다.

핵심 포틀릿은 로그인, 데이터 등록, 데이터 검색 등이고, 응용 포틀릿은 적설/상강 분석 등의 응용 포틀릿이다.

다. Middleware Service

최근의 IT 구조는 "계층화된" 구조를 지니게 되는데, 이것은 각각의 계층이 다음 계층과 분리되어서 모듈화를 이룰 수 있게 한다. 그러므로 소프트웨어 디자인과 유지보수에 있어서 좋은 성능을 낼 수 있는 것이다. GeoGrid에서는 리소스에 관한 계층과 사용자 중간에 "middleware" 계층을 두어 서비스를 제공하고 있다. "middleware" 서비스는 다음과 같은 기능을 제공한다.

(1) Registration

데이터를 등록하려면 메타데이터, 스키마, 데이터 아이템을 등록해야 한다. GeoGrid의 메타데이터는 ADN (Alexandria-DESE-NASA) 메타데이터 스키마를 사용하고, 스키마 등록은 사용자가 등록한 데이터를 등록하게 하여 다양한 스키마에 대한 일관된 접근 방법을 제공하도록 한다. 이 모든 데이터 등록에 대한 것들은 포탈의 등록 포틀릿에서 이루어진다.

(2) Data Integration

등록된 스키마와 데이터 아이템에 대해서 다양한 자원에 대한 통합 기능을 제공한다. 이 기능은 포탈의 MyPage 포틀릿에서 제공한다.

(3) Indexing

메타데이터에 대한 텍스트 인덱싱을 제공한다. 사용자는 포틀릿에서 텍스트 기반의 검색 기능을 이용할 수 있다.

(4) Workflow

데이터 처리를 사용자에게 보이지 않게 처리하는데 Kepler workflow를 도입한다.

(5) Mapping

즉시 처리되는 맵핑은 MyPage 포틀릿의 한 기능으로서 제공된다.

라. Core Software Service

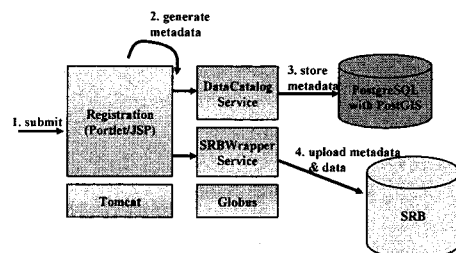
GeoGrid의 System Software Stack은 Globus, Axis, Tomcat, Postgres, Ant, GridSphere, JDK, Samba 등의 프로그램들로 구성된다.

마. Hardware & System Software

각각의 기능을 담당하는 하드웨어들이 한 대 또는 여러대로 구성되어 한 지역에 모여 있을 수도 있고, 분산된 환경으로 멀리 떨어져 있을 수도 있다. 이러한 클러스터를 쉽게 관리해주는 관리 툴도 필요할 것이다.

3.2 데이터 등록 절차

데이터 등록은 사용자가 자신이 보유하고 있는 데이터를 모든 GeoGrid 사용자에게 제공하는 서비스로 커뮤니티 활성화에 큰 도움을 주는 응용 서비스이다. <그림 5>는 데이터 등록의 시스템 흐름 및 응용에 필요한 컴포넌트들을 보여준다.



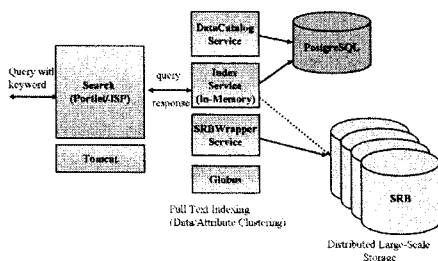
<그림 5> 데이터 등록

먼저 사용자는 자신의 데이터와 자신의 데이터에 해당하는 내용을 Registration Portlet과 JSP를 이용하여 제출을 한다. 후에 Tomcat에 기반하는 Portlet과 JSP는 해당 데이터를 이용하여 메타 데이터를 생성하고, 해당 메타 데이터를 Database에 저장(DataCatalogService)한다. 그리고 대용량 스토리지 응용인 SRB에 SRBWrapperService를 사용하여 생성된 메타 데이터와 사용자가 제공한 데이터를 저장한다. 이는 데이터 검색의 효율을 높이고 대규모 저장 장치를 효과적으로 사용할 수 있게 해준다.

Globus Toolkit 상에서 동작하는 DataCatalogService와 SRBWrapperService는 모두 Grid 서비스이며 Stateless 서비스의 형태로 구현하여 Replication을 이용한 확장성을 고려하여 구현한다. 그리고 Portlet과 JSP로 나누어서 구현되어 있는 웹 응용들을 Portlet과 JSP 중의 하나로 통합하여 추가될 응용에 쉽게 통합할 수 있는 구조로 다시 디자인하고 구현한다.

3.3 데이터 검색

데이터 검색은 사용자들이 등록한 데이터 혹은 GeoGrid이 자체적으로 생성하는 데이터 (이미지 전처리 후의 영상, GeoGrid 상의 application에 의한 결과물)를 효과적으로 검색하여 액세스할 수 있는 핵심 응용 서비스 중의 하나이다. <그림 6>은 GeoGrid의 데이터 검색의 시스템 흐름도와 관련된 시스템 컴포넌트들을 보여준다.



<그림 6> 데이터 검색

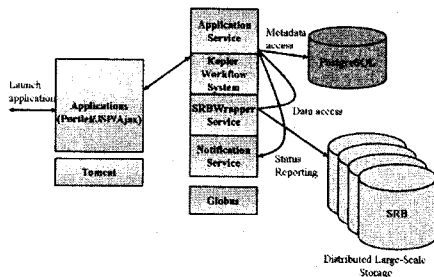
사용자는 GeoGrid Search Page를 이용하여 검색 조건을 시스템에 전송한다. DataCatalogService는 해당 검색 조건이 저장된 메타 데이터에서 일치하는 데이터를 찾아내어 사용자에게 되돌려준다. 이때, 검색 결과에는 분산 대용량 스토리지 상의 위치 정보까지 포함한다. DataCatalogService는 메타 데이터 상에서의 검색을 담당하지만, 메타 데이터에 포함되지 않는 데이터는 IndexService를 사용한다. IndexService는 데이터 아이템에 대한 모든 정보를 Full-Text Indexing을 하고, 주어진 질의와 가장 유사성이 높은 결과를 사용자에게 되돌린다. 즉, GeoGrid의 검색은 2단계로 - 메타 데이터 상의 검색과 Full-Text Index 상에서의 검색- 이루어진다.

사용자는 시스템이 보여주는 검색 결과를 이용하여 데이터에 액세스하거나 SRBWrapperService를 사용하여 해당 데이터를 다운로드 받을 수 있다. 데이터 검색 응용 역시 client(GeoGrid Search Portlet/JSP)와 middleware service(DataCatalogService, IndexService, SRBWrapperService)등이 분리되어 차후 시스템 확장에 효율적으로 대처할 수 있게 한다.

GeoGrid에서 사용하는 데이터에 대한 메타데이터는 ADN metadata schema[reference]를 따른다. ADN은 Alexandria-DL ESE-NASA의 약자로서 여러 가지 목적의 메타데이터 스키마를 정의하는 표준인데, 이중에서 GeoGrid과 관련된 지리 정보를 저장하는 메타데이터 스키마는 Overarching bounding box, Detailed geometries, Elevation, Planet or body, Place-name tied to coordinates, Place-name not tied to coordinates, Event name tied to coordinates, Objects in space, Coordinate system, Vertical and horizontal datum, Projection, Latitudes and longitudes 등의 필드를 포함하고 있다.

3.4 워크플로우 기반 응용프로그램

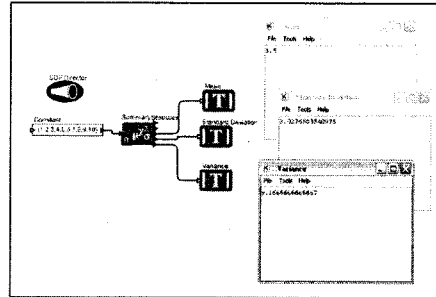
워크플로우 기반 응용 프로그램은 향후 GeoGrid 시스템이 추가될 응용 프로그램 중에서 워크플로우를 기반으로 수행하는 응용 프로그램으로 미국의 GEON의 Lidar 응용이 그 예가 될 수 있다. GeoGrid의 경우는 위성 영상 자료의 전처리부터 워크플로우를 기반으로 할 수 있고 전처리 후의 영상의 재가공도 워크플로우를 기반으로 할 수 있으므로 그 응용 영역이 굉장히 넓다. 워크플로우 기반 수행을 위하여 GeoGrid은 과학 포털에서 가장 널리 쓰이는 Kepler 워크플로우 시스템[reference]을 도입하여 사용한다. GeoGrid의 워크플로우 기반 응용은 <그림7>에서 그 구조와 시스템 컴포넌트들을 볼 수 있다.



<그림 7> 워크플로우 기반 응용

사용자는 GeoGrid application Portlet/JSP/Ajax 등을 통하여 GeoGrid에서 제공하는 응용을 시작할 수 있다. GeoGrid application Portlet/JSP/Ajax은 해당 ApplicationService에 응용의 파라미터 등을 넘겨주고 그 결과를 사용자에게 되돌려준다. ApplicationService는 SRB로부터 데이터를 전송받거나(SRBWrapperService) 해당 응용을 계산 노드에서 실행하는 일들을 Kepler Workflow 시스템의 도움을 받아서 진행을 한다. 그리고 사용자에게 응용의 진행 상황을 NotificationService를 사용하여 E-mail 등으로 통보한다. 이런 간단한 구조는 GeoGrid에 향후 추가될 워크플로우 기반 응용이 쉽게 통합될 수 있게 해준다.

Kepler Workflow System은 과학적인 데이터를 모델링하고 분석하기 위한 응용 프로그램으로써 예는 <그림 8>과 같다.



<그림 8> Kepler Workflow system의 예

Kepler는 실행 가능한 모델링을 만드는 데 비주얼한 표현을 제공함으로써 모델링에 드는 노력을 최소화하도록 해준다.

4. 결론

지구시스템의 이해와 과학적 예측정보 생산을 위한 전 지구 관측 및 예측 모델의 개발과 시스템 구축, 그리고 합리적인 사결정을 지원하기 위한 과학적 정보의 필요성이 대두되고 있다. 우리나라도 기상청을 중심으로 2006년부터 지구관측그룹(GEO)을 통하여 현존하는 상용 통신 인프라를 이용하여 전 지구관측시스템 자료의 수집 및 전달을 위한 구축사업을 하고 있으며, 분야별 국내 관측과 관리 시스템을 운영하고 있다.

국내대학 또한 지구환경변화에 대한 연구는 매우 활발하다. 하지만, 대부분 자료의 수집 및 관리는 개별적으로 운영되고 있으며, 협업 연구 환경 조성이나 자원·자료의 공유를 위한 시스템 개발 등의 전략적 방안이 미비한 실정이다.

본 논문은 각 단체, 기관별로 개별 수집 및 관리하였던 자료들을 통합함으로써 협업 환경에서의 연구 효율성 증대와 관련 연구 분야의 시너지 효과를 높일 수 있는 새로운 지구환경서비스 창출 및 커뮤니티 형성에 기여할 수 있을 것으로 사료된다.

향후 연구 과제는 분석되고 설계되어진 내용을 토대로 GeoGrid 웹 환경을 구축하여 지구환경 관측자료 및 지구관측 위성 영상 자료를 공유·관리하는 일 이외에 이를 분석하고 의견들을 교환하는 커뮤니티 장을 마련하는 웹 포털 서비스 환경을 제공함으로써 지구환경 전반에 걸쳐 선진국 수준으로 도약하게 되는 발판을 마련해야 할 것이다.

참고문헌

- [1] 김지영, 박미경, 오성남, 조하만, 김경렬, “한반도 대기 중 CFC-11과 CFC-12의 배경농도 분포 특성”, 한국기상학회지, 37(1), 65-74, 2001
- [2] 민희경, 김지영, 최병철, 오성남, “Sunphotometer 관측을 통한 서울 지역 대기 중 파장별 에어러솔 광학 깊이의 특성”, 한국기상학회지 38(1), 25-38, 2002
- [3] 조하만, 임영권, 김정우, 김주공, 김지영, “전 지구적 대기 중 CO₂ 농도의 분포 특성”, 한국기상학회지, 36(2), 167-178, 2000
- [4] 환경부, “기후변화 감시, 예측 및 영향평가 기술”, 지구온난화물질 감시 기술 개발 최종보고서, 2002
- [5] Kepler Workflow System
<http://www.kepler-project.org/>
- [6] GeoGrid 분석 및 설계
<http://www.geongrid.org/>