

# Grid-IR 아키텍처 기반의 통합 멀티 검색 서비스 모델 설계

황성빈\*, 김혁호\*, 정용환\*\*, 이필우\*\*, 김양우\*

\*동국대학교 정보통신공학과

\*\*한국과학기술정보연구원

e-mail: {hsb78, hulegea, ywkim}@dongguk.edu, {paul7931, pwlee}@kisti.re.kr

## Design of Unified Multi-Search Service Model based on Grid-IR Architecture

Sungbin Hwang\*, Hyukho Kim\*, Pilwoo Lee\*\*, Yonghwan Jung\*\*,  
Yangwoo Kim\*

\*Dept. of Information and Communication Engineering, Dongguk University

\*\*Korea Institute of Science and Technology Information

### 요 약

그리드는 분산된 이기종 자원을 활용하는 컴퓨팅 패러다임으로 초기 그리드 자원을 활용하기 위한 표준으로 OGSA(Open Grid Service Architecture)가 제안되었다. 글로벌스 툴킷3은 사용자가 OGSA에 표준에 맞게 그리드를 구성할 수 있는 기능을 제공했다. 하지만 그 이후 웹 서비스 표준의 발전에 따라 그에 맞는 새로운 표준인 WSRF(Web Service Resource Framework)가 정의 되었고, 현재 WSRF 표준에 따라서 GT4가 개발되어 발표된 상황이다. 이에 본 논문에서는 기존의 OGSA 표준에 맞게 구축된 그리드 정보검색 서비스들을 WSRF 표준에 맞도록 새로운 표준 인터페이스를 정립하고, OGSA-DAI(Data Access & Integration) 컴포넌트를 통한 데이터 관리 기능을 추가함으로써 새로운 그리드 정보검색 서비스 프로토타입을 정의하고자 한다.

### 1. 서론

최근 그리드[1],[8],[9]를 이용한 정보 검색 시스템에 관한 연구가 활발히 진행되고 있다. 그리드 정보 검색(Grid-IR, Grid Information Retrieval)[4]이란 그리드 컴퓨팅 환경에서의 정보 검색을 의미하며, 기존의 정보 검색 시스템에 그리드의 확장성, 효율성, 그리고 보안 등의 장점을 추가하여 더 나은 정보 검색 서비스를 제공한다.

본 논문의 목적은 글로벌스 툴킷3(Globus Toolkit version3)을 기반으로 구현된 정보 검색 시스템을 바탕으로, 그리드 서비스의 동적 할당 및 재구성할 수 있는 있도록 시스템을 확장하는 것이다. 또한 분산된 데이터 관리 측면에서 OGSA-DAI(Data Access & Integration) 컴포넌트를 기반으로 데이터 접근 및 통합을 위한 표준 인터페이스를 개발함으로써 좀 더 나은 환경의 정보 검색 서비스를 제공하려는 목적이 있다. 그리고 그리드 환경 구축 시 사용되는 대표적인 미들웨어인 글로벌스 툴킷4를 기반으로 새로운 그리드 정보검색 서비스 모델을 제시하려고 한다.

본 논문의 구성은 2장에서 본 연구와 관련되어 활용할 수 있는 기술들에 대해서 알아보고, 3장에서 글로벌스 툴킷4와 OGSA-DAI 컴포넌트를 기반으로 시스템 구축 시 필요한 요구사항 및 서비스 모델 설계에 대해서 설명한다. 그리고 마지막으로 4장에서는 본 논문의 결론과 향후 연

구에 대해서 기술한다.

### 2. 관련 연구

#### 2.1 글로벌스 툴킷

그리드 환경에서 초대형의 거대 문제들을 해결하기 위해서는 그리드로 연결된 유휴 자원들을 찾아내는 자원 검색 서비스, 할당된 작업들의 처리 순서를 결정하여 분산시키는 스케줄링 서비스, 시스템 안정을 위한 그리드 보안 서비스, 컴퓨팅 자원들을 사용할 때 발생하는 비용 처리를 위한 사용자 계정 서비스 등이 필요하다. 이와 같은 서비스들을 제공하기 위해 여러 가지 미들웨어들이 존재하며 그중에서 대표적인 것이 글로벌스 툴킷[2],[3],[10],이다. 현재 글로벌스 툴킷은 버전 4까지 제공되고 있으며, 글로벌스 툴킷의 구성은 크게 5개로 분류할 수 있다.

- Common Runtime Components: Common Runtime Components는 GT4 웹 서비스와 함께 라이브러리 및 툴들로 구성되고, 그것들이 독립적으로 서비스 될 수 있도록 하는 역할을 담당한다.

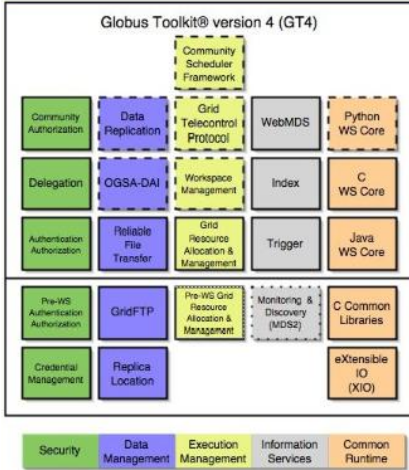
- Security: Security는 GSI를 사용하여 공개키 암호화 방식을 제공하고 그에 따른 보안을 담당한다.

- Data Management: Data Management는 대규모 데이터의 접근, 전송, 그리고 검색할 수 있는 서비스를 제공한다.

- Information Services: Information Services는 MDS로 GT4이전 버전부터 사용되어 왔고, 모니터링부분과 자원발견을 웹 서비스로 제공받을 수 있게 하는 역할을 담당하면서 VO(Virtual Organization)들의 자원을 모니터링하기도 한다.

- Execution Management: Execution Managements는 글로벌스 툴킷이 제공하는 웹 서비스 사용 전후의 실행처리, 모니터링, 작업의 취소를 담당하고 있다.

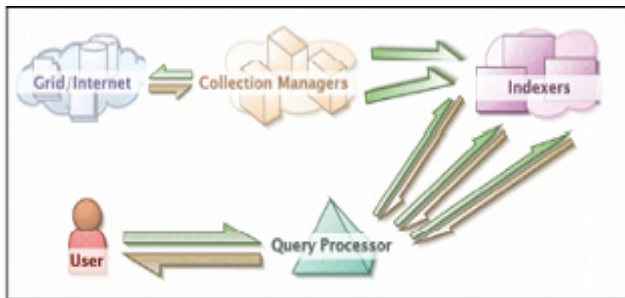
다음과 같은 기능을 수행하는 글로벌스 툴킷4의 구조는 그림 1에서 확인 할 수 있다.



(그림 1) 글로벌스 툴킷4의 구조

2.2 그리드 정보검색(Grid-IR: Grid Information Retrieval)

그리드 정보검색[7]이란 OGSA기반 그리드 환경에서의 정보검색을 말하며, 기존의 정보 검색 시스템에 그리드의 확장성, 효율성, 그리고 보안 등의 장점을 추가하여 더 나은 정보 검색 서비스를 제공한다. 그리드 정보검색의 구조는 일반적인 검색엔진과 비슷한 구조를 가지며, 그림2에서 그리드 정보검색 시스템의 구조를 확인할 수 있다.



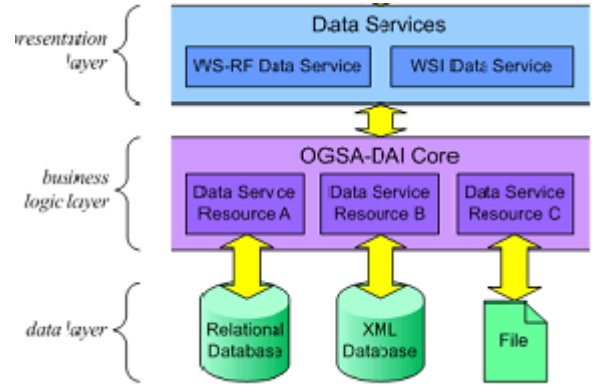
(그림 2) 그리드 정보검색 시스템의 구조

그림 2에서도 볼 수 있듯이, 그리드 정보검색 시스템의 구조는 웹크롤러라 불리는 검색로봇이 문서를 찾으려면 그 문서를 모아서 로컬에 저장하여 관리하는 CM(Collection Manager), 그 문서를 인덱싱 하기 위한 IS(Index/Search) 그리고 검색 결과에서 연관된 부분의 추출과 여러 개의 문서 컬렉션에 질의를 담당하는 QP(Query Processor)의 구조로 되어 있다.

3.4 OGSA-DAI(Open Grid Service Architecture-Data

Access Integration)

OGSA-DAI[5],[6]는 미들웨어 기반구조로 사용되며, 그리드상의 데이터 자원들을 제공하는 웹서비스이다. 그리드 환경에서 이 웹서비스를 제공하면 형태가 다른 데이터 자원들에 대해 사용자는 질의, 업데이트, 전송 등의 기능을 수행 할 수 있다. 또한 데이터의 위치, 데이터의 구조, 데이터 전송, 데이터 통합, 그리고 데이터 분석과 프로세싱 기능들을 제공한다. OGSA-DAI는 서비스 지향 구조(SOA)를 사용하여 데이터 자원에 대한 동일 표준 방식의 접근을 촉진시키기 위해서 OGF의 OGSA 표준과 협력하고 있으며, 이와 더불어 WS-DAI의 표준 구현을 제공하기 위한 연구가 진행 중이다. 초기의 OGSA-DAI 는 OGSi 기반구조상에 기초를 두었으며, 최근 배포판은 웹 서비스 상호운용 조직(WS-I), WS-I+, 그리고 웹 서비스 자원 프레임워크(WSRF)를 지원한다.



(그림 3) OGSA-DAI 구조

3. 통합 멀티 검색 서비스 모델의 설계

3.1 통합 멀티 검색 서비스 모델의 요구사항

본 논문에서 제안한 통합 멀티 검색 시스템 모델은 글로벌스 툴킷4와 OGSA-DAI 컴포넌트를 기반으로 환경을 구축하며 그리드 정보검색 시스템 설계 시 필요한 요구사항은 다음과 같다.

1. 글로벌스 툴킷4 기반의 그리드 정보검색 서비스: Grid-IR 시스템의 구조는 일반적인 정보 검색 시스템과 유사한 구조를 갖고 있으며, 분산된 정보 검색 서비스 즉, CM, IS, QP로 구성된다. 이와 같은 정보 검색 서비스들을 GT4 기반으로 구성된 그리드 환경에서 운영할 수 있는 정보 검색 서비스가 요구된다.

2. Grid-IR 서비스의 연동을 위한 표준 인터페이스: 글로벌스 툴킷4 기반의 정보 검색 서비스들을 연동 및 통합할 수 있는 표준 인터페이스 기술로 웹 2.0을 이용한다. 따라서 사용자가 웹 인터페이스를 통해 작업을 수행할 수 있도록 그리드 정보 검색 서비스들을 연동하는 표준 인터페이스가 필요하다.

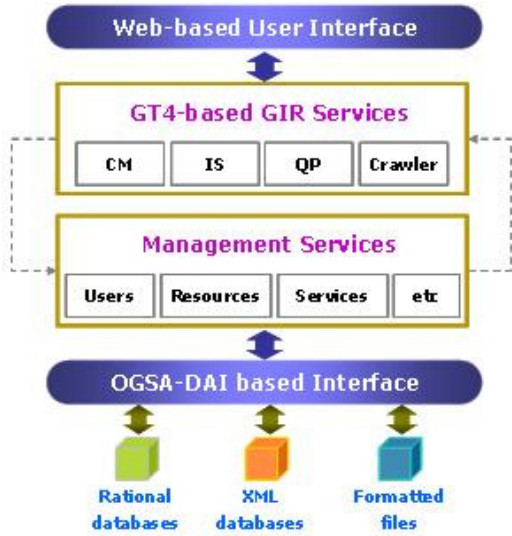
3. 데이터 통합을 위한 OGSA-DAI 기반 표준 인터페이스: 분산된 데이터의 통합 및 관리를 위한 방법으로 OGSA-DAI 컴포넌트를 이용한다. 따라서 OGSA-DAI 컴

포넌트 스펙에 따라 분산된 데이터를 통합할 수 있는 인터페이스를 필요로 한다.

4. 가상조직내의 서브그룹 생성 및 운영: 그리드 컴퓨팅 환경 즉, 가상조직(VO)내에 존재하는 자원들을 활용하는 방안으로 가상조직내의 자원들을 서브그룹으로 재구성하는 기능을 요구한다.

### 3.2 통합 멀티 검색 서비스 모델의 구조

본 논문에 제안한 Grid-IR 아키텍처 기반의 멀티 통합 검색 시스템은 그림 4에서 확인할 수 있듯이 크게 3부분으로 나누어 볼 수 있다.



(그림 4) 통합 멀티 검색 서비스 모델

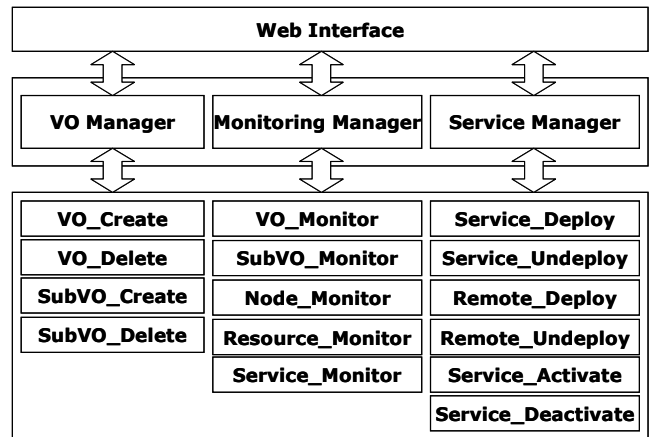
첫째는 데이터 관리를 위한 OGSA-DAI 부분이다. OGSA-DAI 기반의 인터페이스는 지리적으로 분산된 다량의 데이터의 접근 및 통합을 위한 컴포넌트를 제공한다. 따라서 분산된 데이터를 통합 관리 할 수 있다. 둘째는 그리드 서비스 부분으로 정보검색 서비스 부분과 자원 모니터링 서비스 부분으로 나누어 볼 수 있다. 정보 검색 서비스 부분은 Grid-IR 아키텍처에 따라 CM, IS, QP 서비스를 구성하고 Crawler 서비스를 추가하였다. 그래서 실질적인 정보 수집, 데이터 관리, 인덱싱 작업, 그리고 질의 처리를 할 수 있도록 구성된다. 다음으로 자원 모니터링 서비스 부분은 가상조직 모니터링, 가상조직내의 노드 모니터링, 컴퓨팅 자원 모니터링, 그리고 그리드 서비스 모니터링 작업을 수행한다. 마지막으로 세 번째는 가장 상위에 존재하는 웹 기반의 인터페이스 부분이다. 사용자는 웹 기반의 인터페이스를 통해 데이터 관리, 가상조직의 상태 파악, 그리드 서비스 및 자원의 상태 관리, 그리고 필요한 서비스 및 자원의 요청을 수행할 수 있다.

### 3.3 통합 멀티 검색 서비스 모델의 설계

그림 5는 Grid-IR 아키텍처 기반의 통합 멀티 검색 서비스 모델의 핵심 모듈을 도식화한 그림이다. 통합 멀티 검색 서비스의 핵심 모듈은 가상조직 관리자(VO

Manager), 모니터링 관리자(Monitoring Manager), 서비스 관리자(Service Manager)로 크게 3부분으로 나누어 볼 수 있다.

- 가상조직 관리자는 다음과 같은 기능을 수행한다.
  - VO\_Create: 새로운 가상조직을 생성한다.
  - VO\_Delete: 기존에 존재하는 가상조직을 제거한다.
  - SubVO\_Create: 가상조직에 존재하는 그리드 노드를 이용하여 하위 가상조직을 생성한다.
  - SubVO\_Delete: 생성된 하위 가상조직을 제거한다.
- 모니터링 관리자는 다음과 같은 기능을 수행한다.
  - VO\_Monitor: 등록된 가상조직을 모니터링한다.
  - SubVO\_Monitor: 임의의 가상조직에 존재하는 하위 가상조직을 모니터링한다.
  - Node\_Monitor: 가상조직에 존재하는 노드들을 모니터링한다.
  - Resource\_Monitor: 가상조직에 존재하는 자원을 모니터링한다.
  - Service\_Monitor: 가상조직에 존재하는 모든 그리드 서비스를 모니터링한다.
- 서비스 매니저는 다음과 같은 기능을 수행한다.
  - Service\_Deploy/Undeploy: 로컬 머신에 그리드 서비스를 배치/해지한다.
  - Remote\_Deploy/Undeploy: 원격의 노드에 그리드 서비스를 배치/해지한다.
  - Service\_Activate: 하위 가상조직에 필요한 서비스를 구성한다.
  - Service\_Deactivate: 하위 가상조직에 존재하는 서비스들을 해지한다.



(그림 5) 통합 멀티 검색 서비스 모델의 핵심 기능

## 4. 결론 및 향후 연구

본 논문에서는 Grid-IR 아키텍처 기반의 통합 멀티 검색 서비스 모델에 대한 요구사항들을 제시하였다. 또한 OGSA-DAI 기반의 인터페이스를 제공하여 다양한 종류의 데이터들을 지원할 수 있도록 하였다. 이를 통해 분산된 데이터에 대한 접근과 그들을 통합할 수 있는 기능을 제공한다. 그리고 웹 기반의 인터페이스를 통해 사용자가

쉽고 편리하게 서비스를 활용할 수 있도록 하였다. 그리고 기존 GT3 버전의 서비스들을 GT4 기반의 서비스로 새롭게 구성하여 시스템의 성능 및 서비스 활용률을 증대시킬 수 있을 것으로 예상된다. 이와 더불어 다양한 OS와 데이터를 가상조직 기반의 환경으로 구축할 수 있어서 그리드의 확장성, 효율성 그리고 보안성을 확보할 수 있다.

향후 연구로는 본 논문에서 제시한 Grid-IR 아키텍처 기반의 통합 멀티 검색 서비스 모델에 대한 요구사항을 바탕으로 새로운 버전의 정보검색 서비스를 구현해야 하며, OGSA-DAI 컴포넌트를 통한 데이터 관리 요구사항을 상세하게 도출할 필요가 있다. 그리고 데이터 관리를 위한 인터페이스 구현해야 하며, 이를 기반으로 통합된 그리드 정보검색 시스템을 구축하고 테스트베드를 운영하여 실제적인 성능 실험을 수행할 계획으로 있다.

### 참고문헌

- [1] IBM Web site, <http://www-128.ibm.com/developerworks/kr/grid/newto/>
- [2] The Globus Alliance Web site, <http://www.globus.org>
- [3] KISTI Web site, [http://www.ksc.re.kr/infor/infor\\_3\\_b.htm](http://www.ksc.re.kr/infor/infor_3_b.htm)
- [4] 박찬호, 김혁호, 이승하, 김양우, 이필우, "그리드 정보 검색 시스템을 위한 동적 가상조직 설계 및 구현, 한국정보과학회 학술발표 논문집 2005, 2005년 7월
- [5] The OGSA-DAI Project Web Site, <http://www.ogsa-dai.org.uk/documentation/ogsadai-wsi-2.2/doc/background/architecture.html>
- [6] The OGSA DAI Project Web Site, <http://www.ogsadai.org.uk/documentation/ogsadai-wsi-2.2/doc/background/architecture.html>
- [7] Kim, Y.P and SeungHa Lee and Pilwoo Lee and Newby, G.B. "Grid Information Retrieval Management System for Dynamically Reconfigurable Virtual Organization" IEEE 2006
- [8] I. Foster, "What is the Grid? A Three Point Checklist", GRIDToday, July 20, 2002.
- [9] I. Foster and C. Kesselman, The Grid 2: Blueprint for a New Computing Infrastructure, Morgan Kaufmann, 2nd edition, November 18, 2003.
- [10] Ian Foster, A Globus Toolkit Primer (Draft), <http://www.globus.org/>