e-Science 기반 계산화학 교육환경(e-Chem) 설계

안부영[1], 서정현[1], 김지영[1], 조금원[1], 차지영[2]
한국과학기술정보연구원[1], 한남대학교[2]
{ahnyoung, jjery, yes, ckw}@kisti.re.kr[1], fuhaha1231@nate.com[2]

Design of e-Science Gateway for Computational Chemistry

Bu Young Ahn[O1], Jeong Hyun Seo[1], Ji Young Kim[1], Kem Won Cho[1], Ji Young Cha[2]
Korea Institute of Science and Technology Information[1], Hannam University[2]

요 약

요즘 들어 컴퓨터 처리 능력의 향상에 따라 사이버인프라스트럭처(Cyberinfrastructure)를 이용하는 계산과학이 주목을 받고 있다. 그 중에서도 대용량 데이터의 복잡한 계산과 시뮬레이션을 동반하는 계산화학 연구 분야에서의 컴퓨터 활용은 매우 중요하다. 계산화학을 간단하게 설명하자면 컴퓨터를 이용한 계산을 통하여 이론 화학의 문제를 다루는 화학의 한 분야라고 말할 수 있다. 계산화학 분야의 연구를 위하여 고성능 컴퓨터와 데이터를 처리, 분석하는 계산화학 도구는 이론연구자 및 실험연구자 모두에게 있어 필수적인 요소이다. 더불어 계산화학 연구자간의 협업과 원격지에 있는 사이버인프라스트럭처 자원의 활용을 위해 e-Science 환경에서의 연구 및 교육 환경이 개발되어야 한다.

이에, 본 논문에서는 한국과학기술정보연구원(KISTI)이 보유 및 운영하고 있는 사이버인프라스트럭처(고성능 컴퓨터, 초고속 네트워크)를 기반으로 컴퓨터에 익숙하지 않은 계산화학 관련 연구자 및 전공자들이 인터넷 상에서 계산화학 분야 교육을 받을 수 있는 e-Science 기반 계산화학 교육을 위한 환경을 설계하고자 한다. 이를 위해 1) 세계적으로 유명한 GridChem, CICC, NBCR 웹사이트를 이용하여 발표된 논문을 분석하였으며, 2) 분석된 결과를 가지고 주로 사용되는 계산화학 도구의 통계를 산출하여, 3) 이를 바탕으로 KISTI 사이버인프라스트럭처를 활용한 e-Science 기반 계산화학 교육 환경(e-Chem)을 설계하였다.

1. 서 론

계산과학(Computational Science, Scientific Computing)은 과학이나 공학에서 수치적 방법과 컴퓨터 계산을 이용하여 문제를 해결하는 분야이다. 계산과학은 계산과 컴퓨터 그리고 정보처리 자체에 관해 연구하는 전산학과는 구분된다. 또한 계산과학은 기존의 과학과 공학 분야에서 사용되는 방법인 이론 및 실험을 통하여 연구 대상에 대한 이해를 얻어내는 것이 아니라, 주로 컴퓨터를 이용하여 수학적인 모델을 해석하는 방법을 통하여 연구 대상을 이해한다.

계산과학 중에서 계산화학(Computational Chemistry)이란 컴퓨터를 이용한 계산으로 이론화학의 문제를 다루는 화학의 분야 중 하나이다. 복잡계에 속하는 화학 문제는 컴퓨터의 힘을 이용하여야만 풀 수 있는 경우가 많다. 컴퓨터를 이용한 경우 전산화학이라 불리기도 한다. 최근에는 컴퓨터 처리능력 발달에 의해 실험, 이론과 어깨를 나란히 하는 제3의 연구수단이 될 정도로 발전하였으며 다음과 같은 방법을 이용해 화학 문제를 다룬다.

- 분자궤도함수 이론 (MO: Molecular Orbital)
- 분자동역학 (MD: Molecular Dynamics)
- 몬테카를로 방법 (MC: Monte Carlo)
- 분자역학 (MM: Molecular Mechanics)
- 밀도범함수이론 (DFT: Density Functional Theory)

계산화학의 중요성은 화학뿐만 아니라 재료 및 생물학까지 커다란 영향을 미치고 있다. 어떠한 연구이든지 분자 구조와 기능의 이해에 따라 모든 연구결과물의 산출이 달라지기 때문이다. 이를 위해서는 적절한 컴퓨팅 자원(하드웨어, 소프트웨어, 네트워크) 즉, 사이버인프라스트럭처를 이용하는 것이 중요하다. 이런 필요성의 충족은 e-Science 환경으로 해결이 가능하다.

2. 계산화학 도구 조사 분석

사이버인프라스트럭처란 다양한 컴퓨팅 자원과 서비스를 상호 이용 가능하도록 제공하는 과학과 공학 연구에 지원이라는 측면에서 설명된다. 정보기술의 발전은 거대과학 및 공학 사업을 위해 필요한 사이버인프라스트럭처의 개발을 가져왔다. 그러나 대규모 연구 커뮤니티에서는 가능한 사이버인프라스트럭처의 활용이 소규모 개인연구자에게는 접근이 어려운 실정이다.

계산화학분야에서는 커뮤니티, 이론 연구자, 실험연구자에 이르기까지 컴퓨터 자원에 대한 필요성이 증가하고 있기에 사이버인프라스트럭처를 기반으로 구축된 GridChem과 같은 계산화학 커뮤니티는 미국의 연구재단(NSF)에서 상당한 예산을 지속적으로 투자하여 슈퍼컴퓨팅센터의 커다란 구성요소로 자리 잡고 있다.

2.1 GridChem

미국 일리노이대학교 슈퍼컴퓨팅센터(NCSA, National Center for Supercomputing Applications)에서 개발한 그리드기반 계산화학 연구 환경인 GridChem은 사용하기 쉬운 인터페이스로 컴퓨터에 익숙하지 않은 화학 및 생명분야 커뮤니티 연구자들에게 사용 장벽을 낮춤으로써 사이버인프라스트럭처의 활용성을 높였다는 평가를 받고 있다. 이 프로젝트는 미국 연구재단(NSF, National Science Foundation)이 지원하고 있으며, 다른 연구기관과의 연계를 통해 글로벌 사이버인프라스트럭처를 구축하여 연구자들에게 각자의 데스크톱 환경에서 그리드 컴퓨팅을 활용 가능한 서비스를 제공하고 있다[2].

그리드 컴퓨팅 환경이라 함은 보안, 작업 추적, 라이센스 계약, 자원 할당, 대용량 데이터, 사용자 인터페이스를 모두 포함한다. GridChem은 그리드 컴퓨팅에 접근할 수 있는 광범위한 화학분야 커뮤니티에서 활용되고 있다. 또한 high-end 컴퓨팅 활동, 그리드 인프라스트럭처, 협업 도구 등의 활용에 있어 선도적 역할을 수행하고 있다. 그리고 화학분야 연구자의 연구 목적을 기반으로 계산화학 도구 사용을 원할히 제공할 수 있도록 지원하는 컴퓨터 기술자를 포함하고 있다.

GridChem 프로젝트는 화학분야에서의 연구 및 교육의 지원뿐만 아니라 타 분야 커뮤니티에 적용 가능한 기능을 배포할 수 있는 선도적인 사례가 될 것이다. 향후 GridChem 기능들은 전기전자, 생물공학, 의학 등의 분야에서 중요한 신물질을 생산하는 산업에 잠재적으로 영향을 미칠 것으로 예상된다. 특히 나노공학과 의약품 개발에 계산화학이 매우 중요한 분야이기에 국가 경제와 융합 기술의 가용성을 전제로 GridChem의 개발을 지원할 것이다.

GridChem에서는 총 31편의 논문을 조사·분석하였는데 그 중에서 25편의 논문이 Gaussian을 활용하여 논문을 작성한 것으로 나타났다. 나머지 6편의 논문에서는 AMBER, NAMD 등의 도구가 활용되었다.

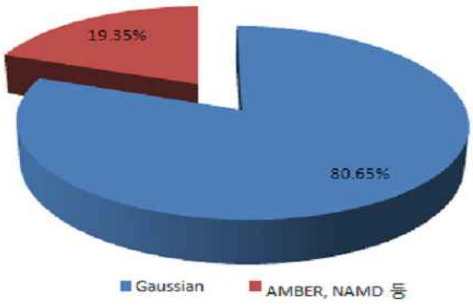

그림 1. GridChem 계산화학도구 활용현황

GridChem은 게재된 논문의 대부분이 Gaussian를 통해 밀도범함수 이론을 계산하였다. 밀도 범함수 이론은 반-경험 양자화학 방법의 하나로 물질, 분자 내부에 전자가 들어있는 모양과 그 에너지를 양자 역학으로 계산하기 위한 이론의 하나이다. 페르미(Fermi) 등이 고안한 것을 월터 콘(Walter Kohn)교수가 발전시킨 것이다. 이를 통해 어떤 분자가 세상에 존재할 수 있는지 없는지의 여부, 특정 분자의 모양과 성질 등을 예측할 수 있다. 컴퓨터를 사용하는 계산화학 중에서 가장 널리 쓰이는 양자역학 계산 분야 중 하나이다. 이 방법은 분자의 크기가 커져도 계산 시간의 증가가 이전에 사용했던 방법보다 훨씬 작으므로 더 큰 분자들을 다룰 수 있게 한다. 또한 얻어진 결과들이 순-이론 방법으로 구한 결과들과 견줄 수 있을 만큼의 우수성을 가지고 있기 때문에 많은 사람들이 사용하고 있다.

2.2 CICC

미국 인디애나대학교에서 개발하여 운영 중인 CICC (Chemical Informatics and Cyberinfrastructure Collaboratory)에서는 GridChem에서와 같이 구체적인 도구를 논문을 통해 확인하기 어려웠다. CICC에서는 총 25개의 논문을 조사하였는데 계산화학보다는 화학정보학(Chemoinformatics)에 관한 내용이 많았다. 이러한 특성 때문에 개방된 다양한 데이터베이스를 구축한 웹사이트의 데이터를 이용하여 예를 들어 PubMed, PubChem, PDB를 이용하여 데이터를 모으고 필요한 데이터를 구분하는 연구가 주로 수행되고 있었다.

CICC는 그리드 기술을 사용하여 분산된 화학적 도구, 시뮬레이션, 문서, 데이터베이스를 생물학적 자원과 편리하게 통합 가능하게 구성되어 있다. CICC는 오픈 웹 서비스 기반 구조를 이용하여 NIH Molecular Libraries Initiative 및 PubChem으로 부터 받아오는 데이터를 통합할 수 있는 인터페이스와 애플리케이션 및 데이터베이스를 개발하여 제공하고 있다[2]. 또한 화학정보학 교실이라는 가상의 교실을 웹사이트에 구축하여 인디애나대학교의 학부생과 대학원생들이 어디서든 바로 접속할 수 있는 사이버 교육의 체계가 잘 갖추어져 있다.

CICC와 같은 화학적인 정보 처리가 가능한 사이트를 통해 신약 개발과 화학적 구조를 컴퓨터를 통해 예측하는 QSAR(Quantitative Structure Activity Relationship)의 학문들이 더 발전할 수 있을 것으로 사료된다.

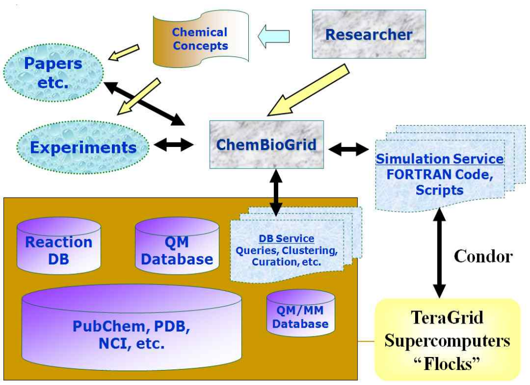

그림 2. 분자모델링을 위한 Varuna 환경[2]

2.3 NBCR

미국 샌디에고대학교에서 개발하고 운영 중인 NBCR (Natioinal Biomedical Computation Resource)은 국가 생물의학자원을 이용하여 거대 생물 의학 연구를 지원하는 사이트이다. NBCR의 주요 목적은 그리드 인프라스트럭처 기반의 계산화학 환경에 모든 연구자들이 접근을 가능하게 한다는 것이다. 물리, 화학, 생명 분야의 연구자들이 데이터와 실험을 통한 각자의 연구를 수행하여 의학적으로 중요한 신약물질 설계, 뇌 화상 진단과 심혈 관계 질환 등을 비교 유전체학로부터의 생물학적 기관에 이르기까지 광범위한 조사의 통합 환경을 제공한다[4].

NBCR에서는 2004년부터 2010년까지 게재되었던 논문 총 319편을 조사하였다. NBCR은 국가적인 생물의학 계산 자원으로 다중 척도의 생물의학 연구가 가능하도록 계산 자원 및 커뮤니티를 구성하여 연구자들을 지원하고 있다. 조사한 논문은 크게 세 부분으로 나눌 수 있었다. 첫 번째, 생물의학계에 응용할 수 있는 다중척도 다양한 시뮬레이션 도구에 관한 부분, 두 번째, 신경과 기관의 생물물리학적인 다중척도 모델링 환경을 조성하는 부분, 마지막으로 이러한 다중척도 생물의학 연구를 위한 유연하고 확장된 사이버인프라스트럭처 프레임워크를 구축하는 부분으로 나누어진다.

NBCR은 GridChem과 다르게 생물학적 분야를 다루다보니 화학 분야보다는 더 복잡하고 다양한 도구들을 사용하게 되었다. 이러한 많은 도구를 세포 이하 과정의 통합 모델링 및 계산(APBS, PDB2PQR, FETK, SMOL, AutoGrow, FERBE, AMBER, CHARMM, NAMD, Gromacs, IMOD, ImageJ), 다중척도 생물 의학 모델링의 시각화 환경 생성(AutoDock, MGLTools), 그리드 컴퓨팅과 다중척도 생물 의학 애플리케이션의 분석(GAMA, Opal, TxBR, GAMESS, QMview)을 위한 도구로 나누어 볼 수 있다. 이 밖에도 통계적인 도구로 ANOVA, MATLAB 등의 소프트웨어가 사용되었고 전자 현미경인 TEM, SEM 등도 세포를 관찰하는데 사용되었다.

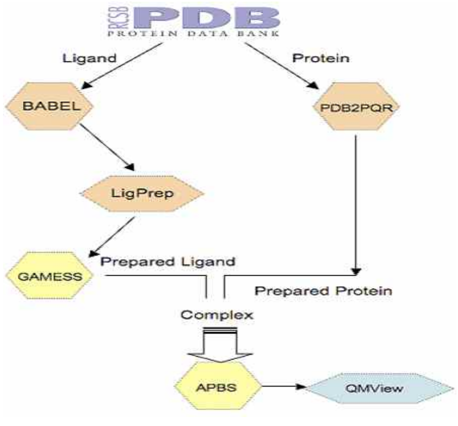

그림 3. NBCR 계산화학도구 활용사례[4]

<그림 4>는 NBCR에서 사용되는 다양한 도구 현황이다. 특히 ANR DOCKING에서는 주로 AutoDock를 사용하였고 ParadisEO라는 프레임워크를 사용한다는 것도 알 수 있었다.

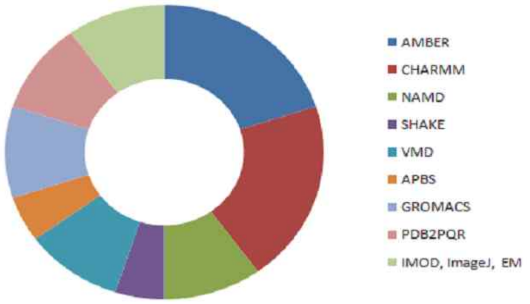

그림 4. NBCR 계산화학 도구 사용현황

3. 계산화학 교육환경(e-Chem) 설계

2장에서 조사한 내용을 가지고 서울대, 서강대 등에서 계산화학을 강의하는 교수와 조교들을 대상으로 면담을 실시하여 e-Science 기반의 계산화학 교육 환경 개발을 위한 요구사항 조사를 실시하였다. 이를 통해 e-Chem의 활용 가능성을 확인하였다. e-Chem의 목적은 컴퓨터에 익숙하지 않은 계산화학 관련 연구자 및 전공자들이 컴퓨터를 이용하여 계산화학 실험을 쉽게 할 수 있도록 사이버 교육환경을 제공하는 것이다.

3.1 계층별 구조

e-Chem의 계층은 <그림 5>와 같이 시스템 계층과 핵심 미들웨어 계층, 서비스 미들웨어 계층, 사용자 서비스를 위한 웹 포털 계층 등 4개로 구성하였다. 시스템 계층은 OS를 비롯한 시스템 소프트웨어를 장착한 하드웨어 노드 또는 슈퍼컴퓨터 리소스를 포함하며, 핵심 미들웨어 계층은 분산 OS, 그리드 미들웨어 등이 포함된다. 계산과학 서비스 미들웨어에는 계산화학을 지원하기 위한 기본적인 소프트웨어 모듈들이 포함되고, 웹 포털 계층을 통하여 사용자에게 서비스하게 된다.

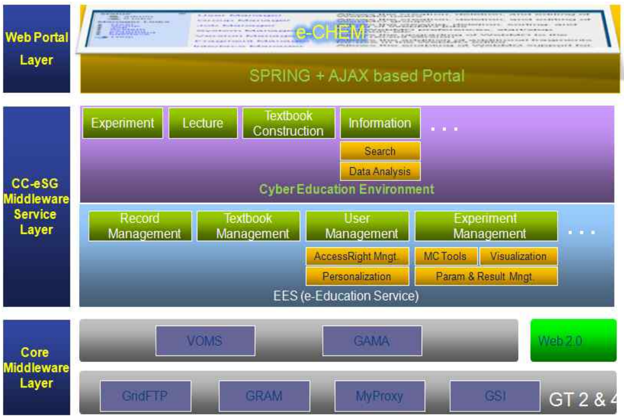

그림 5. e-Chem 계층 구조도

핵심 미들웨어 계층에 논문 분석결과 나타난 연구자들이 주로 사용하는 상용의 계산화학 도구인 Gaussian, AMBER, CHARMM, NAMD 등과 대학에서 계산화학 실습수업에 가장 많이 활용하는 무상의 Gamess를 KISTI가 보유한 클러스터 컴퓨터와 슈퍼컴퓨터에 탑재하여 제공할 예정이다.

3.2 소프트웨어 구성

e-Chem을 구성하고 있는 소프트웨어는 <그림 6>에서 보는 바와 같이 계층별로 클라이언트 티어, 프레젠테이션 티어, 비즈니스 티어, 리소스 티어 등과 같이 4-tier로 구성된다. 각각의 티어는 별도의 서버로 독립될 수도 있고, 경우에 따라서는 하나의 서버 상에 구현될 수 있다.

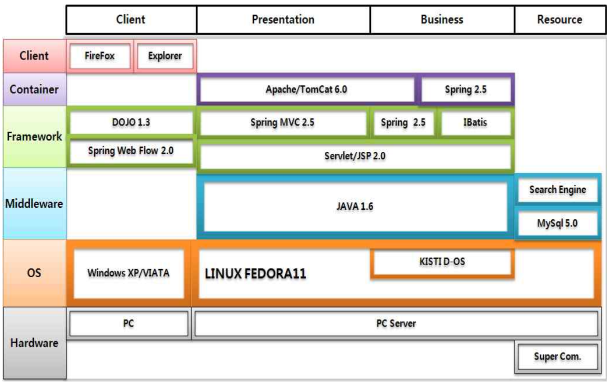

그림 6. 소프트웨어 구성요소

클라이언트 티어에는 FireFox와 MS Explorer를 사용자 인터페이스 도구로 사용하려고 한다. FireFox와 MS Explorer를 채택한 이유는 시장 점유율과 개발 기간 단축을 고려한 효율성 때문이다.

클라이언트 상에서 소프트웨어 개발 도구는 jQuery와 Spring Web Flow를 채택한다. jQuery는 가장 많이 사용되며 잘 만들어진 오픈소스 자바스크립트 라이브러리이다. 분자의 편집과 실험결과를 보여주는 뷰어는 JMOL을 사용하고자 한다. Jmol은 오픈소스 자바기반의 화학구조 3D 뷰어 프로그램이다. 윈도우, 맥OS X, Linux/Unix 기반의 OS에서 데스크탑 기반에서 동작하며 자바 애플릿 기술 적용으로 웹페이지 상에서도 구현이 가능하다.

Spring Web Flow(SWF)는 Spring 프레임워크에서 생성된 모듈이다. 이 모듈은 Spring MVC를 포함하는 Spring의 웹 애플리케이션 개발 스택의 일부이다. Spring Web Flow는 웹 애플리케이션 페이지의 흐름을 관리하기 위한 가장 훌륭한 해결법이 되는 것을 목표로 하고 있으며, 애플리케이션이 애플리케이션 트랜잭션 내에 각각의 단계를 통해 사용자를 가이드하기 위한 마법사처럼 복잡하게 제어되는 탐색(navigations)을 요구할 때 사용하기 위한 강력한 컨트롤러이다.

Apache-Tomcat 6.0은 프레젠테이션 티어와 비즈니스 티어의 컨테이너로 채택된다. 아파치 톰캣(Apache Tomcat)은 아파치 소프트웨어 재단에서 개발된 서블릿 컨테이너(또는 웹 컨테이너)만 있는 웹 애플리케이션 서버이다.

데이터베이스 관리자는 프레임워크가 자동적으로 생성해주는 쿼리문을 분명하게 이해하지를 못한다는 것이다. 그리고 이것이 어떻게 애플리케이션을 유연하게 만드는지 알지 못한다. 애플리케이션의 가장 중요한 병목현상을 가져오는 요인은 데이터베이스이다. 프로그램은 SQL 쿼리문을 넘어서 완벽한 제어를 해야만 한다. 그러면 데이터베이스의 분석이 가능하게 되고 성능을 제대로 관리할 수 있게 된다.

3.3 설계 시나리오

e-Chem은 그리드 개념을 적용하여 KISTI 자원과 해외의 다른 자원도 활용 가능하도록 e-Science 기반으로 <그림 7>과 같이 기본 설계를 하였다.

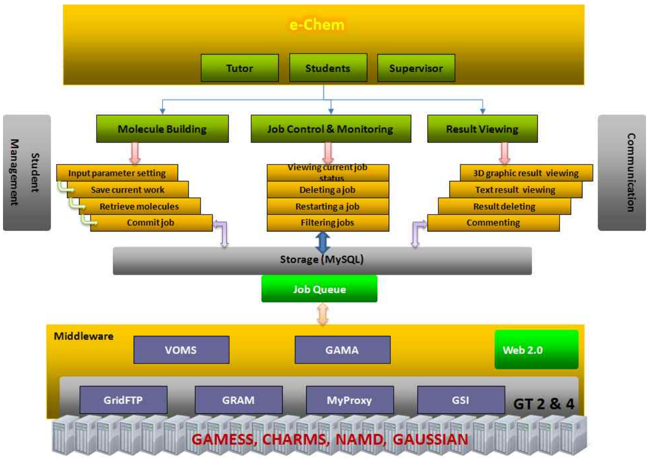

그림 7. e-Chem 기본 설계 구성도

e-Chem의 초기 접속 화면은 웹사이트에 접속하는 모든 사용자들의 초기 화면을 보여준다. 초기 화면으로부터는 로그인, 사용자 등록신청, 공지사항 열람, 사이트 소개 등의 기능을 수행할 수 있다.

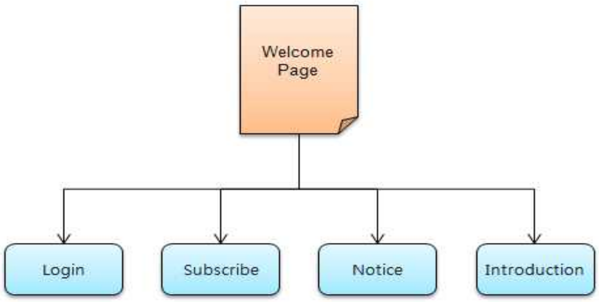

그림 8, 초기화면 구성 모듈

로그인 기능을 위하여 사용자 아이디 입력, 사용자

암호 입력 기능이 있고, 서버로 인증받기 위한 로그인 기능이 존재한다. 공지 사항 열람 기능은 10 줄 정도의 공지 사항의 제목이 표시되며, 스크롤 기능과 페이지 네비게이션 기능을 가지고 있다. 사이트 소개를 위해서는 e-Chem을 설명하는 자료들이 링크되어 있으며, 사용자 등록신청은 강사 수준의 관리자의 등록신청을 위한 것으로 신청 화면으로 연결된다.

e-Chem의 분자편집기능은 실험할 분자 모형을 설계하고 형태를 3D 화면으로 시뮬레이션할 수 있는 기능이다. 하나의 화면에서 파라미터 입력, 원자 위치 수정, 3D 보기 등이 가능하다. 분자 시뮬레이션은 JMOL 애플릿을 사용하며, 입력 파라미터는 Gamess 등의 계산화학도구의 파라미터를 받아들이도록 한다. 분자 편집 화면에서 가능한 모든 액션은 팝업 메뉴를 통하여 사용자에게 제공되며, 저장, 실행, 열기 등의 서버와 직접적으로 동작하는 기능만 제외하고 Client에서 모든 동작이 가능하다.

e-Chem의 분자 시뮬레이션 기능은 입력된 파라미터에 따라서 편집된 분자를 3D로 시뮬레이션하는 기능이다. JMOL 오픈 소스 애플릿을 채택하되 향후 독자적인 개발을 모색하는 방향으로 수행하려고 한다.

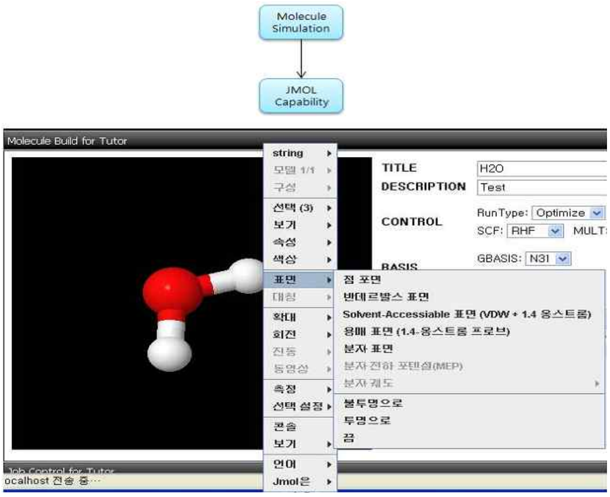

그림 9. JMOL 연동 구성 모듈

e-Chem의 강사 작업 통제 기능은 학생들이 수행한 모든 작업을 모니터링하고 통제할 수 있도록 하는 기능이다. 강사는 자신의 강의에 참여한 학생들이 수행한 모든 작업의 목록을 볼 수 있다. 목록은 학생별 필터링, 강의별 필터링, 작업 상태별 필터링이 가능하다.

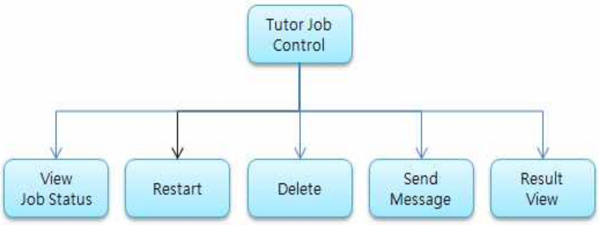

그림 10. 강사 작업통제 모듈

마지막으로 결과보기 기능은 저장된 입력 파라미터를 사용하여 Gamess 등의 계산화학 도구를 수행한 실험 결과를 사용자에게 3D 및 텍스트로 제시하는 기능이다. 설명 결과 표시는 기본적으로 3D로 시뮬레이션 된다. 3D 시뮬레이션에 사용되는 도구는 역시 JMOL이며, JMOL은 향후 독자 개발의 판단 기준이 된다. 텍스트 보기는 사용자의 선택이 있을 때만 새로운 브라우저로 표시되며 결과 텍스트의 수정은 불허한다.

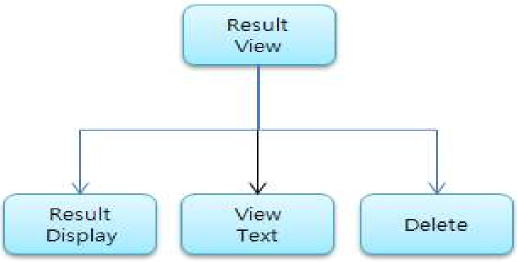

그림 11. 결과보기 구성 모듈

4. 결 론

지금까지 일부의 논문을 통한 계산화학 도구를 조사하여 e-Chem을 개발하기 위한 기본 자료를 살펴보았다. 본 연구에서 수행한 조사는 다소 편협적인 면이 있다고 할 수도 있지만 사이버인프라스트럭처를 이용하는 계산화학이라는 특정 분야에서 주로 사용되는 분석도구(소프트웨어)를 파악할 수 있었다. 이를 통해 KISTI에서 개발하여 제공하여야 할 교육을 위한 e-Science 기반 계산화학분야 교육 환경 포털(e-Chem)이 지향해야 할 점을 미리 엿볼 수 있었다.

이렇게 조사된 내용을 가지고 서울대, 서강대 등에서 계산화학 강의를 하는 교수와 조교들에게 면담을 통한 개발 타당성 및 향후 활용을 위한 요구사항 조사를 실시하였다. 조사된 요구분석 사항을 기반으로 계산화학분야 연구자 및 전공자를 지원하기 위한 e-Science 기반 계산화학분야 교육 환경인 e-Chem의 전체적인 설계를 수행하였다.

본 설계를 바탕으로 KISTI가 보유한 사이버인프라스트럭처를 이용한 시스템이 개발되어 제공된다면 우리나라 계산화학 분야의 연구자 및 전공자들에게 많은 도움을 줄 수 있을 것이라고 사료된다.

[참고문헌]

[1] TeraGrid <http://www.teragrid.org>
[2] GridChem <http://www.gridchem.org>
[3] CICC <http://www.chembiogrid.org>
[4] NBCR <http://www.nbcr.net>