

주요 공학 분야 오픈 소스 소프트웨어들의 고성능 컴퓨팅 활용 동향 분석

윤태호
한국과학기술정보연구원
e-mail:thyoon@kisti.re.kr

Analysis of trends in high performance computing utilization of open source softwares in major engineering fields

Tae Ho Yoon
Korea Institute of Science and Technology Information

요약

본 논문에서는 국가슈퍼컴퓨터 5호기의 혁신지원 프로그램 지원 분석을 통한 5개의 주요 공학 분야를 고성능컴퓨팅 활용도가 높은 오픈 소스 소프트웨어 총 5종과 일대일로 연결시켜 선정하였다. 전 세계 고성능컴퓨팅 인프라 순위를 제시하는 슈퍼컴퓨터 TOP500 전수조사와 선정된 오픈 소스 소프트웨어별 전문가를 통한 집중 자문과 별도의 전문가 발표 그룹 및 참석자들로부터 설문문을 통해 활성화 수준을 분석하였다. 다양한 분석 결과 전체 500대 슈퍼컴퓨터 운영 기관들 공식 홈페이지를 통해서는 약 3.1% 수준만이 주요 공학 분야 오픈소스 소프트웨어의 직접 설치를 지원하는 것으로 조사되었다. 다만 등록되지 않는 고성능 컴퓨팅 인프라 및 산학연관의 특수 사용 분야를 고려해야 하므로 그 수치는 높아질 수 밖에 없다. 제시된 소프트웨어별로 전문가 FSI를 통해 10년이 넘는 장기간에 걸친 버전 업데이트, 튜토리얼, 매뉴얼, 커뮤니티, 컨퍼런스 등이 활발하고 탄탄하게 조직화 되어 이뤄지고 있음이 확인되었다. 소프트웨어 활성화 수준이 높은 단계에 따라 제시된 5종의 오픈 소스 소프트웨어는 이중 컴퓨팅 등 급속한 고성능 컴퓨팅 인프라의 환경에서도 보조를 맞추는 소프트웨어 상용 기술의 신속 탑재의 강점을 지닐 것으로 예상된다. 그 외 동종 분야 및 다른 분야에서의 다양한 오픈 소스 소프트웨어들도 활용성은 꾸준히 증가할 것이라 판단된다.

1. 서론

과학기술 분야 연구자들은 이론, 실험, 컴퓨터 계산의 3가지 방법을 통해 해당 분야의 개발 기술을 확장해 가고 있다. 컴퓨팅 인프라의 지속적 발전으로 점점 컴퓨터 계산을 활용한 기술 개발 효과가 높아감에 따라 해당 계산과학 영역은 점차 독립적인 기술 범위로 확장되어 가고 있다.[1] 또한 한 분야를 넘어 융합적 관점에서 해결되어야 하는 문제도 점차 늘어나는 추세로 계산 요구 수준이 높아가고 있다.

4차산업혁명의 시대 상황 영향에 따른 공학 분야 관심이 늘어가고 있으며 계산과학공학(CSE)으로의 명칭 사용도 자주 눈에 띄고 있다. 그동안 컴퓨터 성능, 계산 요구시간 등에 따라 모델링 형상 단순화, 반응표면분석법(RSM), 촉차모델(ROM) 방법 등 축약된 계산 방식들을 사용하여 효율적인 방법으로 계산 결과들을 활용해 오고 있다.[2] 하지만 평균 8년 이면 전 세계 1위 슈퍼컴퓨터가 500위 밖으로 밀려서 더 이상 슈퍼컴퓨터로 인정받지 못할 만큼 고성능 컴퓨팅 인프라의 발전이 해외 업체들 위주로 빠르게 진행 중이다.[3] 이에 따라 소프트웨어 기술 개발은 컴퓨팅 인프라 요구 성능에 맞춰 상

호 보안을 통해 최적화되도록 신속한 노력을 기울여야 한다.

과학기술 분야 연구자들이 계산과학공학을 이용하기 위한 방법으로는 영리 회사와 같은 기술개발 집단에 의해 폐쇄적으로 개발되는 비용을 지불하고 사용해야 하는 상용 소프트웨어가 있다. 이에 반해 비용이 전혀 없고 개발 코드가 완전하게 공개되는 오픈 소스 소프트웨어를 활용하는 것과 개별 연구실 또는 개인이 자체 코드 개발을 수행하는 것이 큰 틀에서 이뤄지는 계산과학공학 적용 방법이 있다.[4]

오픈 소스 소프트웨어를 활용하는 과학기술 분야 사용자들에게는 빠르게 변화하는 계산과학공학 최신 기술 습득 및 범용적인 전 세계 연구자들과의 커뮤니티 네트워크로 발전되는 실시간 소프트웨어 개발 과정에 동참하는 효과가 클 수 있다. 개별적인 연구실 코드로의 폐쇄성을 갖는 코드 개발도 특정 기술 분야에서의 중요기술 확보로는 유용할 수 있겠으나 기반 라이브러리, 전후처리기 연동 등 수반되어야 하는 노력들이 상대적으로 클 수 있다. 이에 특정한 문제 해결용 코드 개발에서 멈추지 않고 이의 범용성을 갖춘 코드로 확대될 수 있는 오픈 소스 소프트웨어에 연계되도록 하면 코드 안정성, 사용 편의성 등이 더해져 전 세계적으로 기술 활용 효율성이 높

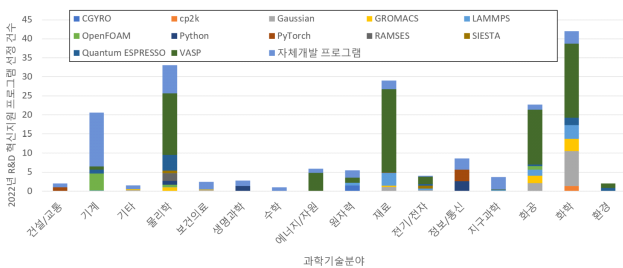
아질 수 있다.

본 논문에서는 과학기술분류체계 내에서의 공학 분야에서 활용되는 상용 소프트웨어 대비 무상의 장점이 있지만 개발 주체의 기술지원 부재 및 검증 한계의 단점을 갖고 있는 오픈 소스 소프트웨어를 대상으로 계산과학공학 활용성을 검토하였다. 과학기술분류체계[5]를 기준으로 하는 주요 공학분야별 활용도가 높은 오픈 소스 소프트웨어를 국가슈퍼컴퓨터 5호기 누리온 지원 프로그램 통계로부터 오픈 소스 소프트웨어 서비스 분석을 통해 선정하였다. 선정된 소프트웨어에 대해 전 세계적으로 고성능 컴퓨팅으로의 주요 활용 수준을 분석하고 국내 오픈 소스 소프트웨어 활용 전문가들을 대상으로 FGI 방법을 통한 심도 있는 활용성 분석이 포함되도록 하였다. 이를 통해 제시된 5종의 오픈 소스 소프트웨어가 앞으로의 이종컴퓨팅 등 고성능 컴퓨팅 발전에 맞춰질수 있는 활용 정도 및 활성화를 예측해 보았다.

2. 주요 공학분야 오픈 소스 소프트웨어 확보

2.1 주요 공학분야 및 오픈 소스 소프트웨어 선정

중점으로 분석하고자 하는 주요 공학분야 고성능컴퓨팅 활용 오픈 소스 소프트웨어를 결정하기 위해 KISTI 국가슈퍼컴퓨팅센터의 연 3회에 걸친 슈퍼컴퓨터 자원을 무상을 나눠 주는 산학연 연구자들을 위한 R&D 혁신지원 프로그램[6]을 통해 검토하였다. 2022년 총 3회에 걸친 R&D 혁신지원 프로그램의 과학기술분야별 전체 선정 건수와 활용 소프트웨어를 [그림 1]에서 제시하고 있다. 고성능컴퓨팅의 활용 빈도가 높은 계산과학공학 분야로는 우선 화학, 물리학, 재료, 기계, 항공 분야의 순서로 선정되고 있음을 알 수 있다. VASP[7]의 상용 소프트웨어와 자체개발 프로그램이 다수를 차지하고 있다. 오픈 소스 소프트웨어로는 CP2K[8], GROMACS[9], LAMMPS[10], OpenFOAM[11], SIESTA[12], Quantum ESPRESSO[13] 등이 주로 활용되고 있다. 공학 분야로 기계, 재료, 항공은 선정 빈도가 높음을 알 수 있으며, 소프트웨어 활용도와 함께 종합적으로 분석한 결과 에너지/자원 및 전기/전자 분야를 대상으로 포함되도록 결정하였다.



[그림 1] 과학기술분야의 2022년 R&D 혁신지원 프로그램 선정건수 및 활용 소프트웨어 분포

2.2 주요 선정 오픈 소스 소프트웨어 소개

GROMACS는 고성능 분자동역학 시뮬레이션 분야에서 가장 많이 사용되는 오픈 소스 소프트웨어 중 하나로 화학/화학공학/재료공학/에너지공학 등 다양한 분야의 계산 연구에 활용되고 있다. 그래픽 처리 장치(GPU)가 시뮬레이션 분야에 본격적으로 사용되어 하이브리드 MD 기술이 등장하면서 GROMACS는 CPU-GPU 이기종 가속 프레임워크를 도입하였고, 시뮬레이션 분야에서 GPU 사용이 주를 이루는 현재는 분자동역학 분야에서 인기있는 코드가 되었다.

LAMMPS는 분자동역학 시뮬레이션에서 전해질 내 확산 속도, 전극-전해질 계면에서 형성되는 전기 이중층 내에서의 반응 등을 계산하기에 적합합니다. 다양한 조건들을 설정 가능하여 목표로 하는 환경을 적합하게 모사할 수 있다. 자체적으로 다양한 적분법과 계산 방법들을 가지고 있어 상황에 따라 적절하게 조정이 가능하다.

OpenFOAM은 주로 유동해석을 수행할 수 있는 유한체적법을 중심으로 하는 편미분방정식 솔버 개발용 클래스 라이브러리와 그것을 이용해 작성된 어플리케이션을 포함하고 있다. 현재는 주로 CPU 기반에서 동작하는 3가지 종류의 소프트웨어로 분기되어 개발 및 배포가 이뤄지고 있으며 주기적인 버전 업데이트 등으로 사용자층이 두텁다.

Quantum Espresso는 DFT, 평면파 기저 및 Pseudopotential (PP)을 기반으로 하는 전자 구조 계산 및 재료 모델링 오픈 소스 소프트웨어이다. 상용 툴인 VASP과 동일하게 범용성이 높은 평면파 기저를 사용한다. 평면파 기저는 단순하게 cutoff energy를 증가시키는 것만으로 기저의 퀄리티를 높일 수 있다. 반면 LCAO(linear combination of atomic orbitals) 기저를 사용하는 제일원리계산의 경우 기저의 퀄리티를 높이기 어렵다. QE는 NCPP, PAW, USPP 등 다양한 PP를 사용하여 계산을 할 수 있기 때문에 원하는 상황에 맞는 계산이 가능하다.

SIESTA는 파동함수를 기술하기 위해 Numerical atomic orbital(NAO) 기반의 기저 함수를 활용하는데 NAO는 특정 원자 주변에 국부화된 영역에서만 정의되도록 설정되어 있다. 시뮬레이션 공간 내에서 전자가 이동하는 전하수송 시뮬레이션을 지원함으로써 소재와 나노소자의 전기적 특성을 연구할 수 있는 오픈 소스 소프트웨어이다. 다른 DFT 코드에 비해 반도체 분야로의 활용에 유용하다.

3. 오픈 소스 소프트웨어의 고성능컴퓨팅 활용성 분석

3.1 Top500 순위 슈퍼컴퓨터 활용 분석

Top500 순위의 슈퍼컴퓨터에 해당 관련 인프라 부서에서의 제시된 오픈소스 소프트웨어의 직접 설치 및

서비스가 이뤄지고 있는지를 확인하기 위해 공식적인 슈퍼컴퓨터 홈페이지를 전수 조사하였다. 주로 대학교에서 운영하는 슈퍼컴퓨터나 관련 분야를 연구하는 연구소에서 운영하는 슈퍼컴퓨터에서 적극적인 설치를 통한 활용 지원이 이뤄지고 있음을 알 수 있었다. 슈퍼컴퓨팅 센터의 자체 설치 지원은 [표 1]에서 보는 것과 같이 OpenFOAM의 경우 가장 많은 21개 슈퍼컴퓨터에서 활용되고 있으며, SIESTA가 가장 적은 4곳에 불과하다. 전체 평균 15개 슈퍼컴퓨터에서 설치되어 3.1% 비율만이 적극적으로 서비스가 이뤄지고 있다. 다만 산학연관의 특수 목적별 슈퍼컴퓨터 운영에 따른 소프트웨어 활용 차이가 있을 것이고, 운영 홈페이지에 따로 공개되지 않을 수도 있어서 전체 슈퍼컴퓨터 활용률에서는 높아질 수 있을 것이다.

[표 1] Top500 슈퍼컴퓨터의 오픈 소스 소프트웨어 자체 설치 지원 현황

오픈 소스 소프트웨어	자체 설치 지원 슈퍼컴퓨터 개수	전체 비율 (%)
GROMACS	19	3.8
LAMMPS	16	3.2
OpenFOAM	21	4.2
Quantum ESPRESSO	17	3.4
SIESTA	4	0.8
평균	15	3.1

3.2 전문가 집단 분석 활용 방법 및 사용자 의견

각 오픈 소스 소프트웨어의 공학분야 연계 전문가 자문을 수행하여 이기종 프로세서 적용, 개발코드 구조, 병렬 성능 등 세부적인 오픈 소스 소프트웨어 활용성질의 응답 방식의 FSI를 검토하였다. 또한 제시된 5종의 오픈 소스 소프트웨어의 활용성을 판단하기 위해 KSC2023 행사에서의 ‘공학분야 오픈 소스 소프트웨어’ 포럼을 구성하여 전문가 발표 및 참가자 현장 설문 조사를 진행하였다. 상용 소프트웨어 사용자 포함 총 참석자 28명은 오픈 소스 소프트웨어를 직접 사용하고 있어 본 포럼에 대한 관심이 높음을 확인하였다. 오픈 소스 소프트웨어의 선택 조건(다중 선택 가능)을 문의한 결과 튜토리얼/예제와 검증, 연구분야 인지도, 사용 사례, 기술지원 여부, 포럼/워크숍, 기술자문/용역 순으로 제시하고 있어 활성화 중요한 요소를 파악할 수 있었다. 또한 오픈 소스 소프트웨어를 제대로 활용하기 위해서는 출연연 등 국가 기관에서의 적극적인 지원 정책이 필요하다는 의견들이 대부분이었다.

국내 산학연관 연구자들의 고성능컴퓨팅 서비스 활

용을 지원하기 위한 국가슈퍼컴퓨터 5호기에서는 현재 제시된 5종의 오픈 소스 소프트웨어를 포함하여 여러 종류의 소프트웨어를 직접 설치 후 서비스하고 있다. 이와 관련하여 [표2]에서 보면 단순 설치 및 일반 활용 이외에 솔버 해결 문의는 상대적으로 적음을 알 수 있다. 이는 오픈 소스 소프트웨어 사용자들은 과학 기술 분야 역량에 더욱 초점이 맞춰져야 하는 솔버의 활용 관점에서 인프라 측에서의 지원으로는 원하는 결과를 확인하기 어려울 것이란 생각에 문의 자체를 하지 않는 것으로 추측된다.

[표 2] 국가슈퍼컴퓨터 누리온 사용자 질의응답 분석 건수(2020~2022)

오픈 소스 소프트웨어	단순 설치	일반 활용	솔버 문제
GROMACS	5	14	2
LAMMPS	21	31	4
OpenFOAM	5	15	7
Quantum ESPRESSO	9	11	0
SIESTA	1	1	1
평균	8.2	14.4	2.8

4. 결론

본 연구에서는 국가슈퍼컴퓨터 5호기의 과제 지원 분석을 통한 주요 공학 분야를 고성능컴퓨팅 활용도가 높은 오픈 소스 소프트웨어 총 5종(GROMACS, LAMMPS, OpenFOAM, SIESTA, Quantum ESPRESSO)을 분석 대상으로 선정하고 활용성을 분석하였다.

전 세계 슈퍼컴퓨터 순위를 제시하는 TOP500 슈퍼컴퓨터의 운영 기관들을 대상으로 공식 홈페이지를 통해서 선정한 소프트웨어들에 대해서는 약 3% 슈퍼컴퓨터에서만 내부적으로 직접 설치를 지원하는 것으로 조사되었다. 다만 산학연관의 특수 목적별 슈퍼컴퓨터 운영에 따른 소프트웨어 활용 차이가 있을 것이고, 운영 홈페이지에 따로 공개를 하지 않을 수 있어 전체 슈퍼컴퓨터 활용률에서는 높아질 수 있을 것이다.

주요 공학 분야별 각 오픈 소스 소프트웨어 전문가 FGI를 통해 공학 분야 문제 해결 범위의 다양성 및 활용 효과, 넓은 사용자층 보유가 이뤄지고 있음을 확인하였다. 이는 제시된 소프트웨어별로 10년이 넘는 장기간에 걸친 버전 업데이트, 튜토리얼, 매뉴얼, 커뮤니티, 컨퍼런스 등이 활발하고 탄탄하게 이뤄지고 있는 점으로 해당 추측이 가능하다.

이를 통해 제시된 5종의 오픈 소스 소프트웨어는 이종 컴퓨팅 등 급속한 고성능 컴퓨팅 인프라의 환경에서도 보조를 맞춰 갈 수 있는 소프트웨어 상용 기술의 신속 탑재에 대한 강점을 지닐 것으로 예상된다. 이종 컴퓨팅 등 급속한 고성능

컴퓨팅 인프라의 환경에서도 보조를 맞춰 갈 수 있는 상용하는 소프트웨어 알고리즘의 신속 탑재의 강점을 지닐 것으로 예상된다.

이에 따라 공학분야의 연구자들에게 과학기술 발전을 주도적으로 이끌어 갈 수 있는 효과적인 수단으로 오픈 소스 소프트웨어 개발 및 활용에 동참하기를 적극 제안하고자 한다.

후기

이 논문은 2023년도 한국과학기술정보연구원 (KISTI)의 기본사업으로 수행된 연구입니다.(과제번호: K-23-L05-C02-S08)

참고문헌

- [1] 윤희준, “계산과학분야의 고성능컴퓨팅 교육을 위한 지식 및 기술요소에 관한 연구”, 성균관대 박사학위논문, P223, 2월, 2019년.
- [2] 윤태호, “공학 시뮬레이션 기반 디지털트윈”, 한국과학기술정보연구원, P101, 5월, 2020년.
- [3] TOP500 순위 성능 분포, <https://www.top500.org/statistics/perfdevel/>
- [4] 윤태호, 서동우, “공학해석용 오픈소스 소프트웨어를 활용한 중소·중견기업의 제조 경쟁력 확보 방안”, 한국기술혁신학회 춘계학술대회 논문집, pp. 617-637, 5월 2015년.
- [5] 국가법령정보센터, <https://www.law.go.kr/행정규칙/국가과학기술표준분류체계/>, 법제처, 2023년.
- [6] 국가슈퍼컴퓨터 5호기 R&D 혁신지원 프로그램, <https://www.ksc.re.kr/jwsc/hsjw/hsjwan>
- [7] VASP 상용 소프트웨어, <https://www.vasp.at/>
- [8] CP2K 오픈 소스 소프트웨어, <https://www.cp2k.org/>
- [9] GROMACS 오픈 소스 소프트웨어, <https://www.gromacs.org>
- [10] LAMMPS 오픈 소스 소프트웨어, <https://www.lammps.org/>
- [11] OpenFOAM 오픈 소스 소프트웨어, <https://www.openfoam.org/>
- [12] SIESTA 오픈 소스 소프트웨어, <https://www.openfoam.org/>
- [13] Quantum ESPRESSO 오픈 소스 소프트웨어, <https://www.quantum-espresso.org/>