웹 기반 시뮬레이션 이력출처 데이터 공유 서비스 설계

정영진¹ · 남덕윤^{2*} · 유진승³ · 이종숙³ · 조금원³

Design of Web based Simulation Provenance Data Sharing Service

Youngjin Jung¹ · Dukyun Nam^{2*} · Jinseung Yu³ · JongSuk Ruth Lee³ · Kumwon Cho³

¹Agency for Defense Development, Daejeon 305-456, Korea

^{2*}Dept. of Supercomputing Technology Development, KISTI, Daejeon 305-806, Korea

³Dept. of Advanced Application Environment Development, KISTI, Daejeon 305-806, Korea

요 약

컴퓨팅 기술의 발달과 네트워크의 확산으로 보다 쉽게 실세계의 다양한 현상들을 수치적으로 분석할 수 있는 웹 기반 시뮬레이션 서비스가 활발히 이용되고 있다. 그러나 대부분의 서비스에서 시뮬레이션 수행 내역이 공유/공개 되지 않아 사용자들이 같은 실험을 반복해서 수행해야하고 데이터나 정보를 공유하기 어렵다. 이 논문에서는 첨단 사이언스 교육 허브(EDISON, EDucation-research Integration Simulation On the Net)의 열유체 분야 시뮬레이션 서비스를 중심으로, 기존에 수행된 시뮬레이션 수행내역을 공개할 수 있는 이력출처 데이터 공유 서비스를 설계한다. 공유 서비스를 위해 EDISON_열유체 시뮬레이션 수행 과정을 “문제 → 계획, 설계 → Mesh → 시뮬레이션 수행 → 시뮬레이션 결과분석 → 보고서”로 정의하고 이에 맞게 기존의 시뮬레이션 수행 정보를 저장한다. 사용자는 이력출처 스토어의 검색/공유 API를 통해 시뮬레이션 수행 내역을 검색하여, 시뮬레이션을 통해 어떤 문제가 어떻게 해결되는지를 이해할 수 있다. 또한 동일한 시뮬레이션을 수행하는 시간과 계산자원의 낭비를 줄일 수 있다.

ABSTRACT

Web based simulation service is actively utilized to computably analyze various kinds of phenomena in real world according to progress of computing technology and spread of Network. However it is hard to share data and information among users on the services, because most of web based simulation services do not share and open simulation processing information and results. In this paper, we design a simulation provenance data sharing service on EDISON_CFD (EDucation-research Integration Simulation On the Net for Computational Fluid Dynamics) to share the calculated simulation performance information. To store and share the simulation processing information, we define the simulation processing step as “Problem → Plan, Design → Mesh → Simulation performance → Result → Report.” Users can understand a problem solving method through a computer simulation by searching the simulation performance information with Search/Share API of the store. Besides, this opened simulation information can reduce the waste of calculation resource to process same simulation jobs.

키워드 : EDISON, 웹 기반 시뮬레이션 서비스, 이력출처, 데이터 공개, 워크플로우

Key word : EDISON, Web based Simulation Service, Provenance, Data open, Workflow

접수일자 : 2014. 02. 13 심사완료일자 : 2014. 02. 19 게재확정일자 : 2014. 03. 03

* **Corresponding Author** Dukyun Nam(E-mail:dynam@kisti.re.kr, Tel:+82-42-869-1043)

Dept. of Supercomputing Technology Development, KISTI, Daejeon 305-806, Korea

Open Access <http://dx.doi.org/10.6109/jkiice.2014.18.5.1128>

print ISSN: 2234-4772 online ISSN: 2288-4165

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서 론

네트워크 및 컴퓨팅 기술의 발달로 실생활에서 일어나는 다양한 현상을 수치적으로 계산하고 분석할 수 있는 웹 시뮬레이션 서비스가 활발히 사용되고 있다. 이 중, EDISON 서비스[1, 2]는 최신의 시뮬레이션 프로그램과 콘텐츠를 서비스 할 수 있는 환경을 제공한다. 특히 고성능의 컴퓨팅 자원을 웹을 통해 쉽게 이용하며 가시화 도구를 통해 결과도 바로 확인할 수 있어 열유체[3, 4], 계산화학[5], 나노물리[6] 분야에서 대학기 천 명이상의 사용자가 활용하고 있다. 많은 사용자가 활발히 사용함에 따라 시뮬레이션 작업 수도 증가하며 이를 수행하는 계산 시간도 증가하고 있다. 그러나 시뮬레이션 수행 내역 및 결과 등을 공유/공개하는 기능이 없어 동일한 시뮬레이션 작업이라도 각 사용자마다 모두 해당 작업의 계산을 수행해야만 결과를 확인 할 수 있는 문제점이 있다.

이 문제를 해결하기 위하여 이 논문에서는 시뮬레이션 수행 내역 중, 계산 시간이 오래 걸리는 작업이나 중복적으로 자주 수행되는 작업을 저장하고 검색하는 이력출처 데이터 공유 서비스를 설계한다. 이 서비스에선 이력출처 저장소(Provenance Store)에 시뮬레이션의 수행내역과 관련정보(문제, 배경지식, 계획, 설계, 레포트, 등)를 저장하고, 데이터 검색 API를 통해 시뮬레이션 수행 이력 정보를 제공한다. 이 논문에서는 주로 EDISON_열유체에서 수행된 시뮬레이션 데이터를 중심으로 공유 서비스를 설계하였다.

II. 관련연구

웹 기반 시뮬레이션 서비스는 게임[7], 모델링[8], 수학/과학[9], 수치분석, 사회학, 데이터흐름 다이어그램, 대수학, 등의 다양한 분야에서 활발히 활용되고 있다. 무엇보다 인터넷을 통해 쉽게 접근할 수 있으며 직관적 인터페이스, 웹 기반의 전처리 및 가시화 도구를 제공하여 사용자들이 별다른 프로그램 설치 없이 서비스를 이용할 수 있는 장점이 있다.

웹 기반 시뮬레이션 응용의 경우, 사용자 편의를 위해 보다 많은 계산자원, 직관적인 인터페이스, 가시화 도구, 등 사용자의 시뮬레이션 수행과 분석을 돕는 서

비스 개선은 많이 이루어지지만, 시뮬레이션 수행 과정과 그 결과를 다른 사용자에게 공개하고 시뮬레이션 결과에 대한 신뢰성을 검토하는 기능들은 대부분 아직 서비스되고 있지 않다.

정보의 라이프사이클 동안 변화를 기록하여 결과에 대한 신뢰성 판단 근거를 제공하는 데이터 이력출처(Provenance)[10-15]는 다양한 정보 (생성, 가공, 결과, 소유주 정보, 가정, 등)를 분석하여 타겟을 좀 더 명확히 이해할 수 있게 돕는다. 이력출처 시스템은 보통 “실시간 사건 수집” → “Time series 데이터 저장소를 지원하는 데이터 모델에 사건을 저장” → “데이터 검색 및 활용” 과정을 거쳐 서비스되며, 마치 기록 저장 소처럼 보여진다[16]. 사건 저장을 위해 이력출처 그래프 모델이 많이 사용되는데 이중 대표적인 것이 Open Provenance Model(OPM)이다. OPM은 객체(entity), 문서(artifacts), 행위자 (actors)를 기본으로 데이터가 처리되는 과정을 기록하며 이를 그래프 형식으로 표현한다[17, 18].

이력출처 기술은 신약임상시험[19], 모바일, 클라우드 서비스, 천문우주, 지구과학, 등의 분야에서 과거 이력정보를 통해 지식을 얻고 협력 연구를 수행하는 등 유용한 정보를 얻는 데 활용된다. 개발 중인 신약의 경우, 다양한 임상시험을 거쳐 그 효과 및 안정성을 검증하는 과정이 필수적이기에, 엄중한 품질 관리 측면에서 법적, 구조적, 과학적인 절차에 맞추어 신약의 효능과 부작용, 등을 기록한다[19]. 이를 위해 다양한 상해 수준에서 참여자(임상의, 연구자, 임상시험 대상자, 보조자, 등)와 각 처리 단계에 대한 진행 과정 (계획, 진행, 평가, 분석), 특이사항, 등을 표준 가이드 라인에 따라 기록한다. 이 기록은 이력출처 저장소에 접근 가능한 사용자가 질의를 통해 검색할 수 있으며, 질의는 데이터의 상세 수준 (속성, 하위 수준 등), 각 데이터간의 관계 (이전단계, 다음단계, 상위), 에 따라 기록된 임상 시험 데이터를 다양하게 검색하고 분석할 수 있다. 검색된 결과는 보통 그래프 형태로 정리되며 각 과정별로 누가 참여했고, 어떤 데이터를 활용하여 결과가 어떻게 나왔는지를 기술한다.

이 논문에서는 시뮬레이션에 대해 이력출처 기술을 활용하여 시뮬레이션 수행 과정 정보를 공개하는 이력출처 데이터 공유 서비스를 설계함으로써, 사용자가 다른 사용자의 시뮬레이션 수행 과정을 검색하고 그 결과

를 검토할 수 있어 사용자가 일일이 모든 시뮬레이션을 수행하는 시간과 노력을 절약할 수 있다.

비슷한 시뮬레이션 작업을 수행하는 것으로 보인다.

III. 웹 기반 시뮬레이션 수행 내역 분석

첨단 사이언스 교육 허브 (EDISON, EDucation-research Integration Simulation On the Net)에서는 연구자들이 개발한 시뮬레이션 프로그램 및 콘텐츠 (보고서, 논문, 시뮬레이션 프로그램 매뉴얼, 등)를 웹을 통해 제공하는 사이버 교육 연구 환경을 제공한다[1, 2].

그림 1. 시뮬레이션 작업 중복 횟수 및 비율
Fig. 1 Simulation job duplication and rate

시뮬레이션이 수행된 내역을 분석하여 그림 1과 같이 중복 비율을 얻었다. 2012년 2학기엔 57~82% 정도의 중복 비율을 보이고 2013년 1학기의 경우, EDISON_Airfoil_Design 프로그램만 약 8%의 중복을 보이고 다른 프로그램은 84~99%의 높은 중복 비율을 보인다. 이는 전체적으로 EDISON 웹 시뮬레이션 서비스가 아직 연구보다는 교육에 보다 집중되었기 때문에

그림 2. 시뮬레이션 프로그램 (2D_comp-2.1, 2D_Incomp-2.1)의 작업 중복 횟수 및 처리 시간
Fig. 2 Simulation job duplication and processing time

그림 2는 2012년 2학기과 2013년 1학기 EDISON_열유체 서비스에서 활용된 시뮬레이션 프로그램 중 가장 중복 횟수가 많은 2D_Comp-2.1과 2D_Incomp-2.1의 중복 횟수와 대기 시간, 수행 시간을 보여준다. 2012년 2학기 2D_Comp-2.1의 경우, 입력 파라메터가 73회 중복된 시뮬레이션 작업들은 36832초 (약 10시간) 정도의 총 계산 시간을 갖고, 1회 중복된 시뮬레이션 작업들은 약 2,064일(day) 정도의 총 계산 시간을 가진다.

이처럼 상당한 비율을 차지하는 파라메터 중복 시뮬레이션 작업 내역을 이력출처 스토어에 저장하고 공유한다면, 계산자원에서 동일한 시뮬레이션 작업을 수행하는 낭비를 줄이고 사용자가 원하는 시뮬레이션 결과를 바로 얻을 수 있다.

그러나 동일한 시뮬레이션 결과를 판단하기 위해서 입력 파라메터와 Mesh 파일이 동일하지 않음이 비교해야 하는데, 모든 시뮬레이션 작업에 대한 Mesh 파일을 비교하는 작업은 많은 비용이 들기 때문에, 공유 서비스를 구축하는데 큰 걸림돌이 될 수 있다. 따라서 그림 3

과 같이 동일 파라미터를 가진 작업에 대한 메쉬 정보를 형상에 따라 분류하여 사용자가 직접 쉽게 검색할 수 있도록 하였다.

그림 3. 공개된 시뮬레이션 작업의 메쉬 파일 선택
Fig. 3 Selecting a mesh type on Provenance store to show the opened simulation job

IV. 이력출처 저장소 (Provenance Store) 구축

그림 4. 시뮬레이션 공유 서비스의 구조
Fig. 4 Structure of simulation sharing service

이력출처 스토어는 정의된 시뮬레이션 수행 과정(그림 5)에 맞게 데이터를 저장한다. 저장된 이력정보는 시뮬레이션 데이터 관리 모듈과 API를 통해 일반 사용자

와 공유되고 검색된다. 이때 모든 시뮬레이션 수행 내역을 공유하기 보다는 많은 시간 계산이 필요한 시뮬레이션 작업이나, 중복이 많이 되는 작업들을 중심으로 공유하면 계산자원의 낭비를 줄여 시스템의 활용도를 높일 수 있을 뿐만 아니라 시뮬레이션 결과를 얻는 시간을 줄일 수 있다. 또한 실생활에서 접하는 여러 문제들[2]에 대한 시뮬레이션 수행 내역을 공개할 경우, 비슷한 문제를 푸는 연구 및 교육에 활용 할 수 있는 장점이 있다.

그림 5. 시뮬레이션 공유를 위한 시뮬레이션 처리 과정 설계
Fig. 5 Design of simulation processing steps for data sharing

그림 5에서 정의된 EDISON 시뮬레이션 수행 과정 (문제 → 계획, 설계 → Mesh → 시뮬레이션 수행 → 시뮬레이션 결과분석 → 보고서)은 사용자가 공유된 시뮬레이션 수행 내역을 차근차근 이해하기 용이하다. 이 중, “Mesh”, “시뮬레이션 수행 정보”, “시뮬레이션 결과”는 EDISON 서비스에서 자동으로 추출할 수 있지만, “문제”, “계획, 설계”, “시뮬레이션 결과분석”, “보고서” 부분은 사용자가 임의로 추가하여 넣어야 한다.

그림 5의 수행 내역 형태로 데이터를 저장하기 위하여 그림 6과 같이 시뮬레이션 프로그램 수행에 대한 과정을 그래프를 작성했다. 이 그래프의 처리 과정에 따라 EDISON 시뮬레이션 데이터가 저장되며 정해진 데이터 수행과정(워크플로우)에 맞게 추출한 시뮬레이션 데이터를 저장한다. 저장된 데이터는 데이터 간의 관계나 각 단계별 속성 값을 중심으로 검색할 수 있다.

그림 6. 시뮬레이션 프로그램 수행에 대한 그래프
Fig. 6 Simulation program execution graph

V. 이력 출처 데이터 활용

시뮬레이션 이력출처 저장소에 저장된 시뮬레이션 수행내역을 API를 활용하여 검색할 수 있다. 그림 8은 이력출처 검색 API를 활용하여 시뮬레이션 수행내역을 검색한 결과의 예를 보여준다.

그림 7. 시뮬레이션 리스트 및 정보 검색 API
Fig. 7 API for simulation list and information

< 공개된 시뮬레이션 수행 내역 예 >

그림 8. 이력출처 스토어에서 공개된 시뮬레이션 작업
Fig. 8 An opened simulation job of Provenance store

그림 8은 그림 5에서 정의된 6단계를 중심으로 사용자에게 공개된 시뮬레이션 수행 내역을 각 탭마다 분류하여 보여준다. 그림 8에서는 “EDISON_CFD를 이용한 에어포일의 두께에 따른 임계 마하수 비교 연구”[20]를

중심으로 문제와, 계획, Mesh, 시뮬레이션 수행 및 결과를 정리하였다. 특히, “계획, 설계” 단계에 저장된 “임계마하수에 대한 에어포일 두께의 영향”과 같은 배경 이론과 NACA0012, NACA0016에 대한 CFD 실험 조건, 그리고 발표된 논문은 수행된 시뮬레이션이 계획에 따라 어떻게 설정되고 수행되었는지를 명확하게 알 수 있다. 이처럼 단순히 시뮬레이션 수행 내역만 공개하기 보다는 시뮬레이션 실험에 필요한 배경지식, 목적, 등도 추가하여 제공함으로써 시뮬레이션에 대한 이해를 돕고 문제를 해결하는 방법을 배울 수 있다.

또한 공개된 시뮬레이션 수행 정보는 사용자가 시뮬레이션 서비스를 효과적으로 활용하는데 유용하다. 예를 들어 사용자가 기 수행된 시뮬레이션 정보를 토대로 시뮬레이션 계획을 세워 시행착오를 줄이는데 도움을 줄 수 있을 것이다.

VI. 결 론

웹 기반 시뮬레이션 서비스에서의 데이터 공유를 지원하기 위하여 이력출처 저장소를 설계하고 시뮬레이션 수행내역과 함께 문제, 배경지식, 계획, 설계, 레포트, 등의 관련정보를 같이 제공하는 공유 서비스를 제안하였다. 이를 통해 사용자들이 배경지식과 함께 시뮬레이션을 통해 문제를 푸는 방법과 결과를 정리하는 방법, 등을 배울 수 있다. 향후 공유된 데이터의 효과를 분석하여 시뮬레이션 서비스의 효율을 높일 것이다.

감사의 글

이 논문은 2013년도 정부 (미래창조과학부)의 재원으로 한국연구재단 첨단사이언스·교육허브 개발사업의 지원을 받아 수행된 연구임 (No. NRF-2011-0020576).

REFERENCES

[1] JongSuk Ruth Lee, Kyu Jin Kim, Jongbae Moon, Dukyun Nam, Jae Hyoung Lim, Jerry Hyeon Seo, Bu-Young Ahn,

Youngjin Jung, Hyoungwoo Park, and Kumwon Cho, "The Status of Higher Education in Science & Engineering in Korea", *e-learning week 2010*, p.107, 2010.

[2] EDUcation-research Integration through Simulation On the Net (EDISON), <http://www.edison.re.kr>

[3] Young Jin Jung, Jongbae Moon, Du-Seok Jin, Bu-Young Ahn, Jerry Hyeon Seo, Hoon Ryu, Ok-Hwan Byeon, JongSuk Ruth Lee, "Web Simulation Service Improvement on EDISON_CFD", *CST 2012*, pp.119-124, 2012.

[4] EDISON_CFD, <http://cfed.edison.re.kr>

[5] EDISON_Chem, <http://chem.edison.re.kr>

[6] EDISON_Nanophysics, <http://nano.edison.re.kr>

[7] AgentSheets, <http://www.agentsheets.com>

[8] AnyLogic, <http://www.anylogic.com>

[9] Explore Learning, <http://www.explorelearning.com>

[10] Provenance, <http://en.wikipedia.org/wiki/Provenance>

[11] Sudha Ram McClelland, "DATA PROVENANCE," Department of MIS, University of Arizona, Tucson, AZ, URL: <http://adrg.eller.arizona.edu>, October 13, 2006.

[12] Peter Buneman, Wang-Chiew Tan, "Provenance in Databases: Past, Current, Future", *SIGMOD Tutorial 2007*.

[13] Peter Buneman, James Cheney, Stijn Vansummeren, "On the expressiveness of implicit provenance in query and update languages," *ACM Transactions on Database Systems (TODS)* 33(4), 2008.

[14] Todd J. Green, Grigoris Karvounarakis, Val Tannen, "Provenance semirings," *Symposium on Principles of Database Systems*, pp.31-40, 2007.

[15] Divesh Srivastava, "Intensional Associations Between Data and Metadata," *SIGMOD*, pp.401-412, 2007.

[16] Liu, Q., Bai, Q., Giugni, S., Williamson, D., Taylor, J., "Data provenance and data management in eScience," *Springer*, 2013.

[17] The OPM Provenance Model (OPM), <http://openprovenance.org/>

[18] L. Moreau, et al., "The Open Provenance Model," University of Southampton 2007. <http://eprints.ecs.soton.ac.uk/14979/1/opm.pdf>

[19] Translational Research and Patient Safety in Europe, <http://www.transformproject.eu/>

[20] Jaeho Lee, Daeyeon Lee, Soo Hyung Park, "Study on Critical Mach Number According To Airfoil Thickness Using EDISON_CFD", 1st EDISON_CFD competition, April 12, 2012.

정영진(Youngjin Jung)

2007년 : 충북대 전자계산학과 박사
2007년 ~ 2010년 : Uni. of MAINE 연구원
2010년 ~ 2014년 : KISTI 선임연구원
2014년 ~ 현재 : 국방과학연구소 선임연구원
※관심분야 : 센서 데이터 융합, 웹 기반 시뮬레이션 서비스 개선, S/W 신뢰성

남덕윤(Dukyun Nam)

1999년 포항공과대학교 컴퓨터공학 공학사
2001년 KAIST 정보통신공학 공학석사
2006년 KAIST 정보통신공학 공학박사
2004년 ~ 현재 : 한국과학기술정보연구원 선임연구원
※관심분야 : 분산시스템, HPC 시뮬레이션 환경, 고장감내 등

유진승(Jinseung Yu)

1995년 전자계산학 공학사
1997년 컴퓨터공학과 공학석사
2001년 ~ 2004년 : 시큐아이닷컴
2004년 ~ 현재 : 한국과학기술정보연구원 슈퍼컴퓨팅연구소 선임연구원
※관심분야 : 컴퓨터 네트워크, 정보 시스템 보안, 분산처리 시스템, 사이버러닝 등

이종숙 (Lee, JongSuk Ruth)

2001 : Ph.D. in Computer Science, University of Canterbury, NZ
2002 ~ : National Institute of Supercomputing and Networking, KISTI (Head, Principal Researcher)
2005 ~ : University of Science & Technology of Korea (Adjunct Faculty)
1998 ~ 2002: University of Canterbury, NZ (Researcher)
1992 ~ 1993: Korea Electronics and Telecommunications Research Institute (Researcher)
※Research Interests: Parallel Computing, Distributed simulation, Simulation based cyber learning, Grid Computing

조금원 (Kumwon Cho)

2000년 한국과학기술원 기계공학(항공우주) (박사)
2001년 ~ 현재 한국과학기술정보연구원 국가슈퍼컴퓨팅연구소 슈퍼컴퓨팅융합연구센터장
2011년 ~ 현재 첨단 사이언스 교육·허브 개발사업(EDISON) 중앙센터장
※관심분야 : CFD, 분산컴퓨팅