Requirements Engineering in Science: Case Study in Physics Software Designing

Lecheng Wang
July 18
CS846 Course Presentation
I. Introduction

Physics Software: design and analysis of approaches for solving mathematical problems in physics domain

Gaussian03: Build molecular model and compute energies

VASP: Optimizing the metal surface and simulating surface dynamics

I. Introduction

Different types of Physics Software

• Personal or intra-group
  Example: batch processing scripts

• Open sourced or published
  Example: MoRiBs

• Commercial
  Example: Gaussian and VASP

This project focuses on this type
I. Introduction

Unfortunately, Requirement Engineering (RE) are ignored in most physics software

- Collaborating various disciplines including mathematics, physics, chemistry, and computer science
- RE is considered to time-consuming in Scientific Computing

However, RE in physics software is necessary. RE helps to understand:
- The physics problem to solve
- Desired functions
- Behavior of a software system
I. Introduction

Challenges

• Requirements change frequently
  Possibly a major reason for missing RE.
  Suggesting RE should include requirements evolution

• It is important to manage requirements for collaboration and reuse
  Solving physics problem requires knowledge sharing
  Communication gaps exists among various working groups and stakeholders

• How to motivate physics software developer to learn and apply RE methods?
I. Introduction

Tasks in currently presented project

Presented in Part II. Theory:
A survey about RE in scientific computing software development.

Presented in Part III. Case Study in Physics Software Designing
Comparing the RE tasks in a real case to the RE model discussed in Part II.
NOTE: software developed before the survey presented in Part II.

Presented in Part IV. Results and Discussion
Could the RE model further improve the RE in the real case studied in Part III?
Is the RE model working well when applying to the RE in the case studied in Part III?
II. Theory – RE model in Scientific Computing

Domain specific requirements model:

RE in scientific computing requires domain knowledge

Domain Specific Requirement Model: High-Level

This high-level model also corresponds to the challenge:
“It is important to manage requirements for collaboration and reuse”
II. Theory – RE model in Scientific Computing

Scientific Knowledge Modeling
• A science problem is first defined
• Mathematical models are created to represent the scientific problem
• Numerical methods are applied to solve the problems
• Mathematical models and numerical methods are created based on scientific assumptions

Y. Li and co-workers, 2011 Seventh IEEE International Conference on e-Science Workshop
II. Theory – RE model in Scientific Computing

Requirements Modeling: incorporating concepts in scientific computing domain

especially important for super computing applications
desirable properties. Example: a fast and accurate simulation for water molecules moving at melting point
external libraries and tools

[Diagram]

other constrains such as development language

Y. Li and co-workers, 2011 Seventh IEEE International Conference on e-Science Workshop
II. Theory – RE model in Scientific Computing

Requirement in Requirements Modeling

- **Data Flow**
  - I/O: input ‘melting point’, and output ‘moving pattern of molecules’

- **Process**
  - functions of processing data

- **Performance**
  - speed and accuracy of computing software

- **Data Definition**
  - ‘meta-data’ of data such as type, format and accuracy

Y. Li and co-workers, 2011 Seventh IEEE International Conference on e-Science Workshop
II. Theory – RE model in Scientific Computing

Integration: how ‘scientific knowledge model’ corresponds to ‘requirement model’
III. Case Study in Physics Software Designing

MoRiBS-PIMC: A program to simulate molecular rotors in bosonic solvents using path-integral Monte Carlo

Tao Zeng a, d, Nicholas Blinov b, Grégoire Guillon a, Hui Li c, Kevin P. Bishop a, Pierre-Nicholas Roy a, d

MoRiBs: to simulate molecular dynamics at low temperature using randomized algorithm


No knowledge about ‘RE Model’ acquired during designing and development. But...

III. Case Study in Physics Software Designing

RE in the development of MoRiBs:

Functional Requirement:
simulate the molecule dynamics at low temperature

Non-functional Requirement:
Accurate, Memory-economical, Restartable, Parallelable (discovered after executing initial version)

Stakeholders:
Customers: Scholars inside and outside UW chemical physics group
Developer and maintainer: Students and postdocs developing the software
Manager: Principal Investigator
?: Peel reviewer for the publication of the software (special in physics software)

Iceberg Slides about RE. D. Berry, University of Waterloo
III. Case Study in Physics Software Designing

Simulate the molecular dynamics at low temperature

The quantum mechanics formula governing molecular moving

Molecules moving according to quantum mechanic formula

Randomized algorithm, statistical algorithm

Y. Li and co-workers, 2011 Seventh IEEE International Conference on e-Science Workshop
III. Case Study in Physics Software Designing

The work station in Chemical Physics Group

An accurate, restartable, and parallelable program to simulate the molecule dynamics at low temperature

Using C++, including Fortran language subroutines

GNU library

Text interface: executing status, statistical averages, ...

Y. Li and co-workers, 2011 Seventh IEEE International Conference on e-Science Workshop
III. Case Study in Physics Software Designing

Input: initial guess, molecular interaction energy, ...
Output: statistical averages

Computing statistical averages by randomized algorithm, based on interaction energy

Accurate, restartable

N/A in studied case

Y. Li and co-workers, 2011 Seventh IEEE International Conference on e-Science Workshop
IV. Results and Discussion

General findings:

Even **without** the knowledge of RE model, most of the components in the RE model have their counterpart in the development of MoRiBs.

• The RE in MoRiBs is efficient and complete
• The RE model is applicable and helpful in real scientific software development
IV. Results and Discussion

How can the RE model improve the development of MoRiB: discovering the requirement of parallel computing earlier

work station in Chemical Physics Group: 3.2GHz CPU, 8GB Memory
Computing should be finished in 1 month even for large scaled system
OpenMP interface supporting parallel computing

Not fast enough
Huge memory enough to support parallel tasks

Y. Li and co-workers, 2011 Seventh IEEE International Conference on e-Science Workshop
IV. Results and Discussion

How to improve RE model to work better when developing MoRiBs?

The requirement of ‘parallelable’ is discovered only after executed.

Considering the ‘peer review’ when publishing the open-sourced scientific software in RE model.
V. Conclusion

Even without the knowledge of RE model, most of the components in the RE model have their counterpart in the development of MoRiBs: both RE model and the RE in MoRiBs are efficient.

The RE model can further improve the RE in the development of MoRiBs

The RE model can be further improved considering the RE in the development of MoRiBs.
Thank You!

Questions and Comments?