Advancing Earth and Environmental Sciences through Workflow-Driven Ontologies

Leonardo Salayandia
PhD Student in Computer Science

Advisors: Dr. Ann Gates and Dr. Paulo Pinheiro da Silva

CyberSHARE Center of Excellence
http://trust.utep.edu/wdo
University of Texas at El Paso

Abstract

Cyber-Infrastructures (CIs) and portals such as GEON [1] and CEON [2] provide services and access to data in support of scientific efforts. With the increasing availability of CI resources, there is also a growing need for mechanisms to integrate data that have varying accuracy and sensitivity and that come from distributed sources. Workflow-Driven Ontology (WDO) is an approach that allows scientists to capture discipline knowledge in the form of ontologies. Subsequently, such discipline knowledge is leveraged towards automatically extracting process knowledge about the operational procedures that scientists use to create models and other scientific artifacts. This process knowledge represents recipes that can be shared among a scientific community to describe how CI resources are integrated to create models and other scientific artifacts.

The goal of this work is to create scientist-centered tools that allow scientists to retrieve and compose CI resources in a way that is consistent with their disciplines and that facilitates the generation of scientific results.

Significance

According to the NSF Blue-Ribbon report entitled “Revolutionizing Science and Engineering Through Cyberinfrastructure” [3], a new age in scientific and engineering research has dawned. There is a need to pursue research in new ways and with increased efficacy. Pushed by computing, information, and communication technology and pulled by expanding complexity, scope, and scale of today’s challenges, CI represents a novel environment on which national and global scientific and engineering research priorities can be addressed.

The complexity of CI technologies can prevent non-technical users from taking full advantage of the technology. The significance of this work lies in creating intuitive tools in support of CI that can be customized to address scientists’ needs. Hence, the ultimate goal of this work is to help scientists take full advantage of CI.

Features that worked well in WDO-II

Features that need improvement in WDO-II

Overall, attitudes toward the WDO-II tool were extremely positive. The features of the tool that were lauded included the graphical representation of process knowledge generated by the tool, its ease of use, and its potential for cross-disciplinary use. The main problem reported included the expressivity limitations of model-based workflows (MBWs). In particular, the process knowledge generated by the tool does not support process iteration and process abstraction (the ability to expand and contract particular steps in the process knowledge to show or hide details).

Next steps include the refinement of the MBW specification to address the reported limitations. To guide such efforts, additional use cases will be gathered in the areas of earth and environmental sciences as they relate to the research projects undertaken by the CyberSHARE Center [4].

Methodology

The WDO-II tool guides scientists through two iterative phases of the WDO approach.

Capturing discipline knowledge

- Brainstorming
  - Identify concepts from the discipline area
  - Ground ontology in the Data, Method, Product classification proposed by Earth scientists

- Relationship Elicitation
  - Identify relationships between concepts
  - Utilize scientific workflow use cases to create the ontology
  - Define input, output, and other relationships between Data and Method concepts

Extracting process knowledge

- Steps to extract process knowledge:
  1. Scientists: Choose a Product concept for which to generate a workflow
  2. Reasoner algorithm: extract process knowledge from WDO

- Tool
  - Displays process knowledge as Model-Based Workflow (MBW)
  - Encodes MBW in OWL
  - Expresses MBW in graphical notation
  - Describes MBW in terminology defined by the scientist

Ontology development is a continuous process. This is reflected in the WDO approach by the iteration of the two phases described above. Ontology refinement is effectively guided by the analysis of the workflow graph. In addition, the tool supports concept harvesting from existing ontologies.

Results

A preliminary usability study was conducted to evaluate the WDO approach and the WDO-II tool in particular. The study was conducted through the Summer Southwest Regional Cyberinfrastructure Workshop held in August 10, 2007 at the University of Texas at El Paso. The workshop invited a group of scientists that included students, faculty, and professionals. The main objective of the workshop was to promote awareness of the role of CI as it relates to scientific research. Workshop participants were presented with a series of tutorials, including a WDO-II tutorial with hands-on exercises. Participants were subsequently asked to assess the tool through a questionnaire and through their contribution in a focus-group meeting.

Conclusions

WDO help scientists capture process knowledge. The benefits of capturing process knowledge is two-fold:

1. It provides a mechanism by which domain experts can document operational knowledge.
2. It provides a machine-interpretable specification of process knowledge that can be leveraged to harness CI resources.

Preliminary results show that the WDO approach and the WDO-II tool are useful for scientists and has potential for cross-disciplinary use. Having tools that can be used across disciplines is a critical requirement to promote use of CI.

References