Building Accessible Tools for Expressing Semantic Workflows

Bharat Bojja1, Katherine Chastain2,3, John Erickson2,3, Deborah McGuinness2,3
RPI Residential Summer Research Experience (Computer Science)
1. Neuqua Valley High School, 2. Rensselaer Polytechnic Institute, 3. Tetherless World Constellation

Abstract

Organizations that sponsor scientific research have begun to require evidence that applicants are strictly applying the scientific method to ensure robust and unbiased experimental design, methodology, analysis, interpretation, and reporting of results. Researchers are being asked to transparently report experimental details so others may more easily and accurately reproduce and extend findings. Provenance metadata that captures the conceptual structure of the analytical process can help accomplish these objectives; the semantics of scientific workflow recipes or plans can be described so that peers can understand methods and reproduce the intended results.

In this work, we have developed a tool that makes it easier for researchers to translate their workflows from flowcharts and textual descriptions into knowledge graphs that capture the conceptual structure and content of their analytical processes. Our Semantic Workflow Builder is a significant contribution to the Tetherless World Constructions growing collection of techniques and tools that enable scientific workflows to be more clearly and consistently shared, improve understanding of all analysis aspects and enable greater reuse and reproduction.

Methodology

1. An example workflow analyzing data from the Collaborative Perinatal Project (CPP) was provided. The CPP Workflow was a complex MATLAB script that analyzed data, produced intermediate datasets, and generated many visualizations.
2. RDF Turtle was originally used to depict the structure and semantics of the CPP workflow. The Resource Description Framework (RDF) is a standard model used to represent relationships between and about entities on the Web.
3. The RDF was converted into an HTML+RDFa document, producing a web page visually representing the structure of the CPP workflow recipe. RDFa lists a set of HTML attributes to enlarge visual data along with machine-readable hints to reduce redundancy and improve accessibility.

Research Objectives

• To create a user-friendly tool that more easily captures and represents the conceptual structure of complex data analysis workflows
• To create a system that can allow researchers to quickly convert their workflow recipes or plans into semantic representations (knowledge graphs)
• To enable researchers to more easily investigate and validate the origins of analytical results, especially the details of the analytical workflows that produced them

Introduction

• The Semantic Web is the next step in connecting all the data sources on the Internet. It is a mesh of information that helps efficiently express relationships between these sources in the World Wide Web.
• Provenance Metadata
  • Refers to sources of information involved in producing or delivering a result
  • Critical in determining whether information is reliable and trustworthy
• The Semantic Workflow Builder is an intuitive way to capture the conceptual structure of scientific workflows.
• Using an intuitive block-oriented programming metaphor, the Semantic Workflow Builder enables the researcher to represent the entire data analysis process:
  • data sourcing → data processing → visualization of results
• The Builder makes it very easy for data analysts to organize and document their process. The Semantic Web-ready metadata it generates enables others to ask questions about the origins of the data and the programs that were used in drawing conclusions.
• The process of searching for relevant answers to questions about the research methodology is efficient and intuitive.
• Ensures that research results are based on the correct data, have been processed using the right methods, and makes it easier for colleagues to reproduce those results and reuse analytical methods.

Results

• Introducing the Semantic Workflow Builder
  • Enables the user to model the conceptual structure of any scientific workflow
  • Once the scientific workflow recipe is captured, the builder automatically generates RDF serialized as JSON-LD, which may then be consumed by a wide variety of Web-based applications and services, including scientific knowledge graphs made accessible via SPARQL endpoints
• Critical workflow metadata is easily readable by both humans and machines. The process of building these workflows is much more natural.

Conclusions

• The Semantic Workflow Builder not only produces machine-readable output, but the code is also easily understandable by humans, allowing for the Web to be more “intelligent”
• It bridges the gap between scientific researchers and their audience.
• This sets the stage for future advancements in the Semantic Web, which can foster better more insightful communication between billions of interconnected devices as they continue to build on to the Internet.

Next Steps

• Future versions would be more adaptable to the user’s needs
• Develop tools to automatically generate workflows by reading existing codebases
• Expand the type of data stored in the workflow to allow more complex queries
• Automate a bigger part of the process by incorporating machine learning techniques within the fields of natural language processing and knowledge representation.

References/Acknowledgements


Press to convert a scientific workflow into a sharable RDF compliant recipe.

Methodology

4. Create and testing this workflow with queries on a cloud-based database called Dydra.
5. Users can obtain information about the workflows by asking questions such as:
   i. What datasets were used in the workflow?
   ii. How many subprograms are there?
   iii. How many of the visualizations rely on a certain dataset?
6. Use Blockly to develop the framework and components to all users to quickly and easily build customized workflows that are more machine-readable and human-readable.
7. Build, Test, and Deploy.