

Semantic Web Service for Bioinformatics Provenance

Research Paper for ICTCC '08

Athman Bouguettaya, Mark Hepburn, Qing Liu, Kai Xu, and Ji Zhang

1. Introduction

There is significant progress being made towards building Semantic Web Services platforms to support bioinformatics, with myGrid [1] and Bio-Moby [2] being the two prominent examples. However, all these efforts are focusing on integrating services from large research and computation centres and making them easily accessible for smaller labs and groups. The management of *provenance*—which refers to the evolution of information (i.e., from where and how the information is obtained)—is still in very early stage in these systems. The benefits of scientific provenance management are not only helping to determine the data's value, accuracy and authorship, but also to enable reproducibility and facilitate users in interpreting and understanding results which can be more important than the results themselves.

There are two other important benefits of provenance management. The first one is for *self-referencing*. Bioinformatic analysis is a long and complex process, and each step usually has a large number of parameters that creates a large space of possible processes. Without a detailed record, it is difficult to remember which of the many trial-and-error paths was conducted. The second benefit is *sharing*. A recorded process can be shared with colleagues and within the research community, which is valuable for others working on similar problems. While there are existing efforts to support process sharing such as myExperiment [3], it requires the users to generate the process manually, which is time consuming and error prone.

2. Semantic Web Service for Bioinformatics Provenance

The Genome Tracker project at ICT Centre Hobart is developing a semantic web services-based system for bioinformatics. One of the focuses of the system is to provide better provenance management for self-referencing and process sharing. The architecture shown in Figure 1 illustrates how the provenance management is embedded into the five layers of the system:

Databases This layer contains all the data sources. The “Log Database” provides the raw data for provenance management.

Web Services In this layer, all the data sources (from the Database layer) and tools are wrapped up as web services, which include provenance management.

Ontologies In this layer, semantic description is added using a bioinformatics (domain) ontology.

Service Manager All the functional components in this layer includes provenance information in their input and feed its results as new provenance information back to the provenance service.

User Interface Here user interactively composes a workflow with the assistance of the stored provenance information, and this process is captured and stored as new provenance information.

Currently, a simple ontology has been created using WSMO [4] to provide semantic description of the included services. We use WSMX [5] for service discovery, composition, and execution. A user interface has also been implemented to facilitate interactive workflow composition.

3. Conclusions

In this paper we presented a semantic web service-based system designed to support the provenance management for bioinformatics research. We described how it is implemented in the five layers of the system to provide better provenance management than existing systems.

References

- [1] Luc Moreau, Terry Payne, and Martin Szomszor, “Using semantic web technology to automate data integration in grid and web service architectures”, in *Proceedings of Fifth IEEE International Symposium on Cluster Computing and the Grid*, 2005, pp. 189–195.

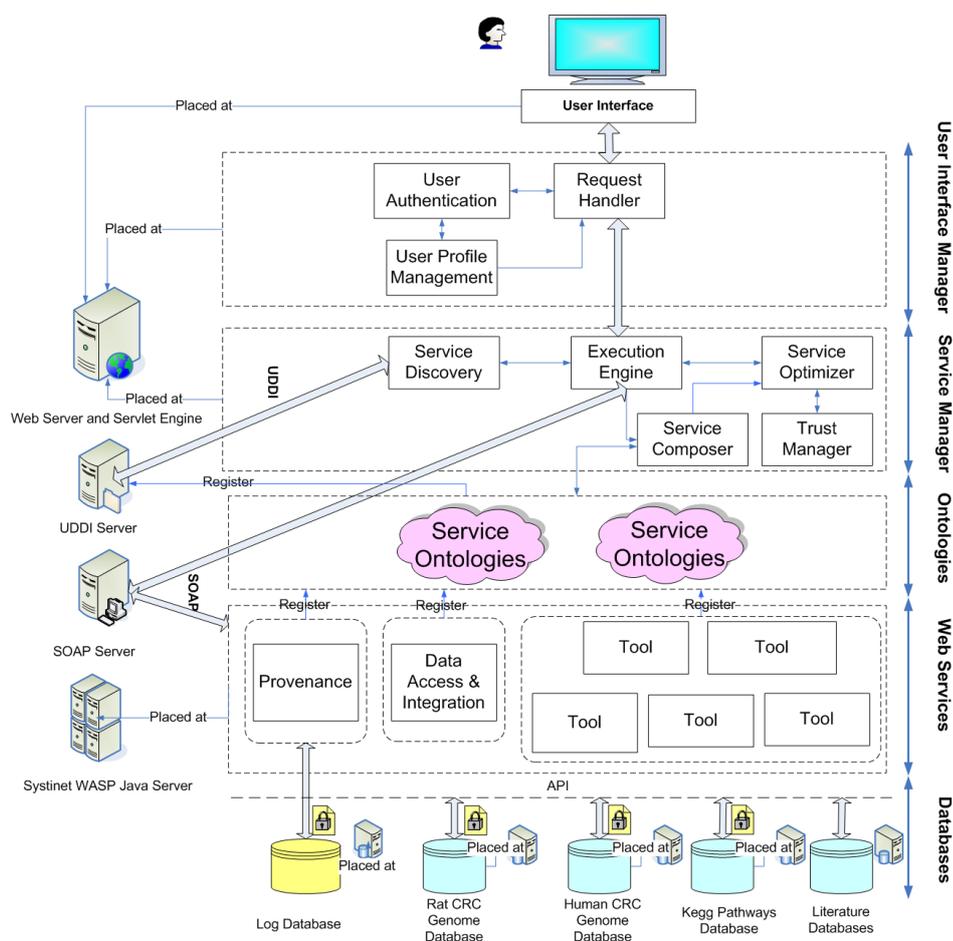


Figure 1: System architecture

- [2] Mark D. Wilkinson and Matthew Links, "BioMOBY: An open source biological web services proposal", *Briefings in Bioinformatics*, vol. 3, no. 4, pp. 331–341, 2002.
- [3] Carole Anne Goble and David Charles De Roure, "myexperiment: social networking for workflow-using e-scientists", in *Proceedings of the 2nd workshop on Workflows in support of large-scale science*, New York, NY, USA, 2007, pp. 1–2, ACM.
- [4] Dumitru Roman, Uwe Keller, Holger Lausen, Jos de Bruijn, Rubn Lara, Michael Stollberg, Axel Polleres, Cristina Feier, Christoph Bussler, and Dieter Fensel, "Web service modeling ontology", *Applied Ontology*, vol. 1, no. 1, pp. 77–106, 2005.
- [5] Armin Haller, Emilia Cimpian, Adrian Mocan, Eyal Oren, and Christoph Bussler, "Wsmx - a semantic service-oriented architecture", in *Proceedings of the IEEE International Conference on Web Services*, Washington, DC, USA, 2005, pp. 321–328, IEEE Computer Society.



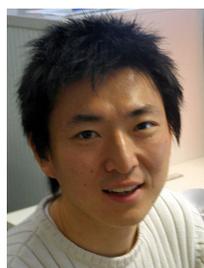
Athman Bouguettaya



Mark Hepburn



Qing Liu



Kai Xu



Ji Zhang