LWGeoWfMS: A Lightweight Geo-Workflow Management System

Ying Du\textsuperscript{1,a}, Pu Cheng\textsuperscript{2,b}

\textsuperscript{1} School of Computer and Information Engineering of Henan University, Kaifeng, 475004, China
\textsuperscript{2} School of Software of Henan University, Kaifeng, 475004, China

\textsuperscript{a} hd_duying@163.com, \textsuperscript{b} chengpul@163.com

Abstract. The disadvantage of the existing workflow systems for spatial information processing flow management is need to extend the special resource managements for the data, models and operations in geographic information systems. It increases the difficulty of development for the systems, the complexity of the architectures and the demand of hardware and software resources. To solve the problem, a light-weight Geo-workflow management system is proposed in this paper. This system has the ability to manage the resource of Geography-related fields. This solution provides a resource sharing platform for different users, designs a script-based process engine for resolving and running the workflows quickly and achieves the functions of process monitoring and traceability at the same time.

1 Introduction

With the rapid development of computer technology and the gradual application of GIS, spatial information processing, which needs to be managed and controlled, become more and more complicated. Geo-Workflow is the workflow technology which can be used to express, manage, control and execute spatial information processing in many works [1]. The theoretical frameworks and technical solutions of Geo-Workflow have been applied in the field of Geographic Model Integration [2, 4], Land Resource Management [3], Spatial Information Processing Services [5] and so on.

Unlike the classical business process and scientific process, the process object of spatial information processing is spatial data with various types and large quantities, and space operations and spatial analysis models for spatial information processing require complex algorithms and intensive computations with many parameters and changing types. Moreover, the existing scientific workflow systems depend on DCIs (Distributed Computing Infrastructures), which contain a lot of components such as data resources, metadata directories, authentication and authorization methods, repositories [6]. These components need to specifically extend to support geospatial information. That greatly increase the difficulties of process management in spatial information processing. Obviously, the existing scientific workflow systems are not easily applied to spatial information processing.

In this paper, we propose a lightweight Geo-workflow management system (LWGeoWfMS), which can manage and control the spatial information processing and is able to complete the tasks in geosciences by scheduling different types of resources. The remainder of this paper is organized as follows. Section 2 describes the reference model of LWGeoWfMS. The design and implementation of
LWGeoWFMS are described in Section 3 and Section 4 respectively. Finally this paper is concluded in Section 5 with plans for future research.

2 Reference Model of LWGeoWFMS

Based on the Workflow Reference Model published by Workflow Management Coalition (WfMC) in 1994[7], we design the reference model of LWGeoWFMS, as shown in Figure 1.

This reference model of LWGeoWFMS has six important interfaces as follows:
- Interfaces 1—Import and export of process definitions
- Interfaces 2—Interaction with the client application of Geo-workflow
- Interfaces 3—Scheduling the execution resources
- Interfaces 4—Collaboration between different workflow management systems
- Interfaces 5—Management and monitoring functions
- Interfaces 6—Scheduling the data resources

The workflow execution service is the core of the whole system, including one or more Geo-workflow engines and resource management engines. The Geo-workflow engine provides the running environment for the spatial data processing such as interpreting process definitions, creating process instances, controlling process routes, changing or recording operational states and communicating with other engines. The process definition tool is used for process modeling which generate a formal description of a spatial data processing. This formal description can be passed to the Geo-workflow engine for analysis and processing through the interfaces 1. The resource management engine is used to schedule various data resources, execution resources and user resources through the interface 6, interface 3 and interface 2 respectively. While the system is running, the interface 2 provides an interaction channel between the system and users by calling the client application program, and the workflow execution service invoke external execution resources through interface 3 to accomplish a specific task. The workflow execution service also needs to interact with other workflow execution services via interface 4, which can make these systems interoperable. The management and monitoring tools perform monitoring management tasks of the process of the workflow execution through the interface 5.

3 Design of LWGeoWFMS

According to the guidance of the reference model, the architecture of LWGeoWFMS system is designed in the three parts: execution server, client application and monitoring server, as shown in Figure 2.

There are four types of users in the LWGeoWFMS: resource contributor, process model designer, task executor and system monitor. The resource contributors register their private resources in the system for the workflow instance to invoke. The process model designers define the workflow model in the visualization environment provided by the LWGeoWFMS system client application, and then they specify the default task executors and resources for the task when the system is executed. The task executors perform the assigned tasks in the workflow instance and the system monitors perform the
monitoring functions while the workflow instance is running.

The execution server is the core of the whole system with the functions of resource management, process management and related scheduling. It consists of two components: the resource management engine and the Geo-workflow engine. The resource management engine is an important part of LWGeoWfMS. It includes the functions of resource registration, resource retrieval and resource scheduling. The resource contributors encapsulate their private resources as common resources and register them in the system through the registration function provided by the resource management engine. If the resources have been registered successfully, they can be retrieved and used in the process of scheduling. In order to organize, manage and schedule different types of resources, different resource templates based on the Resource Description Framework (RDF) are designed. The description of the resource is generated by the resource metadata at the time of registration.

The Geo-workflow engine implemented in Python language includes process analysis, execution and scheduling. According to the rule defined by the process model, the workflow engine resolves a process instance into a list of tasks and then informs the related task executor to carry out his tasks. Python scripts are small, flexible, and easy to scale, so the Geo-workflow engine can be easily deployed in a distributed environment combined with resource management engine, and can encapsulate itself as a resource model automatically.

The client application of LWGeoWfMS provides the users with a customizable visualization interface, which is not only a customized platform for process models, but also the display and operation interface of worklists. The Monitoring Server has the functions of management monitoring, provenance about the entire system.

4 Implementation of LWGeoWfMS
The resource management engine and the client application of LWGeoWFMS are developed on the Visual Studio 2012 platform, using the SQL Server 2008 as the database management system (DBMS) and Microsoft Silverlight which is a rich client Web rendering technology and can give users some novel experiences.

The monitoring server is implemented in C++ and the SQL Server 2008. Figure 3 shows the resource registration interface, through which the resource contributors register their private resources to the system for the workflow instance to invoke. Figure 4 shows the resource management interface. The interfaces of process design and process representation are showed in Figure 5 and Figure 6, respectively.

Fig3 the Interface of Resources Registration  Fig4 the Interface of Resources Management

Fig5 the Interface of Process Design
5 Conclusions
In this paper, the LWGeoWFMS not only implements the basic components of the workflow reference model, but also provide the resource sharing platform for users. Moreover, it uses the Python language to implement the workflow engine and execute the process in an efficient interpreted way. But the research of this paper has many shortcomings: (1) workflow execution server is a centralized implementation, which does not take into account the process of distributed expression and only consider the implementation of distributed resources; (2) while using RDF for resource description, a resource description template has been developed for conforming to the standard only, and many types of resources have not to be further summarized; (3) the interoperation between the implementation and other heterogeneous workflow systems has not been realized. These are the next step to improve in the LWGeoWFMS system.

References
[1] Weske M, Vossen G, Medeiros C B, et al. Workflow Management in Geoprocessing Applications[C]// Proceedings of the 6th ACM international symposium on Advances in geographic information systems. ACM, 1998: 88-93.
[2] Hofer B, Mäss S, Brauner J, et al. Towards a knowledge base to support geoprocessing workflow development[J]. International Journal of Geographical Information Science, 2016:1-23.
[3] Wang Y. Research on the Typical GIS Service Chain Application in Land Resource Management [M]// Geo-Informatics in Resource Management and Sustainable Ecosystem. Springer Berlin Heidelberg, 2015.
[4] Yue P, Zhang M, Tan Z. A Geoprocessing Workflow System for Environmental Monitoring and Integrated Modelling[J]. Environmental Modelling & Software, 2015, 69: 128-140.
[5] Thakker S, Abdul J, B. Potdar M. GeoProcessing Workflow Models for Distributed Processing Frameworks[J]. International Journal of Computer Applications, 2015, 113(1): 33-38.
[6] Terstyanszky G, Kukla T, Kiss T, et al. Enabling scientific workflow sharing through coarse-grained interoperability[J]. Future Generation Computer Systems, 2014, 37(7):46-59.
[7] WfMC, The Workflow reference model 1994.