KNOWLEDGE REPRESENTATION OF REMOTE SENSING QUANTITATIVE RETRIEVAL MODELS

Jingzun Zhang[1,3,4], Yong Xue[1,2*], Jing Dong[5], Jia Liu[1,3], Longli Liu[1,3], Sahithi Siva[2], Jie Guang[1,3]

[1]Key Laboratory of Digital Earth Science, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100094, China
[2] Faculty of Life Sciences and Computing, London Metropolitan University, 166-220 Holloway Road, London N7 8DB, UK
[3] University of the Chinese Academy of Sciences, Beijing 100049, China
[4]Beijing Key Laboratory of Information Service Engineering, Beijing Union University, Beijing 100101
[5]China Institute of Water Resources and Hydropower Research, Beijing 100038, China

ABSTRACT

A large number of quantitative retrieval models have been proposed in recent years, and there is continuous momentum in proposing new ones. Building a model, from design through to implementation stages, involves a process of knowledge collection, organization and transmission. In this paper we introduce the SECI model to manage the conversion of qualitative remote sensing knowledge and propose a mode of knowledge representation on the basis of the ontology for geospatial modeling. We develop a platform based on the above research and demonstrate the efficiency of the knowledge representation mode using this platform.

Index Terms—explicit knowledge, tacit knowledge, quantitative retrieval models, workflow, Ontology, Remote Sensing

1. INTRODUCTION

There have been a number of quantitative retrieval models in recent years and new ones continue to be proposed. Those models represent the knowledge of domain experts on how to produce an application-specific product from available raw data sources [1]. However, some of these are merely a repeat of the existing work (especially pre-processing models), engendering redundancy which goes against the original spirit of sharing. Therefore, the problem of model reuse should be addressed urgently.

Development of a model, from building to implementation stage, can be regarded as a process of knowledge collection, organization and transmission. A reasonable knowledge representation framework is very conductive for researchers to promote development in quantitative remote sensing via making improvement or making up deficiencies. Knowledge building systems are needed urgently to automate the path from data to information and knowledge [2]. Nonaka's theory of organizational knowledge creation, centering on the Socialization, Externalization, Combination and Internalization (SECI) model is probably the most widely cited theory in knowledge management. This paper proposes RS-SECI (Remote Sensing-SECI) model based on a general SECI model focusing on knowledge representation of remote sensing domain.

2. QUANTITATIVE RETRIEVAL MODELS

Quantitative remote sensing retrieval models take the raw data as input and use a set of processes to extract information and knowledge step-by-step. Each process consists of inputs, outputs and operations. The concepts and relationships among these components are specified in ontology. Each process may have two types of inputs representing the initial state: conditional inputs (specifying the preconditions and the relationship conditions between the inputs and the operation, such as data format and spatial reference system) and data inputs (specifying the actual data required by the operation, such as data URL) [3]. Operations use certain algorithms to transform input data into required result as the process's output. The form of outputs usually is intermediate data. These step-by-step processes can be seen as workflows, simple to complex, on the software platform.. Figure 1 shows an original method named SRAP (the Synergetic Retrieval of Aerosol Properties) for aerosol optical depth retrieval[4][5].The data processing procedure follows specific requirements. We set raw data the standard MODIS products of level-1B and aerosol. Obviously, the input data of SRAP model should be processed via information extraction, resize, cut, geometric correction (georef), image mosaic, cloud mask, as shown in Table 1.

3. ADAPTION OF SECI MODEL

According to previous research, knowledge consists of tacit knowledge and explicit knowledge, which differs in the means of expression and transmission. The philosopher Polanyi has put forward the concept of tacit knowledge, followed by a detailed description.[6][7]. Elizabeth Smith analyzed the role of tacit and explicit knowledge in the workplace and proposed ways to recognize, use, share, acquire, teach and measure tacit and explicit knowledge [8]. Koskinent et al. addressed the method of tacit knowledge acquisition and sharing in a project work context [9]. SECI model describes the possible interactions, which exist between tacit and explicit knowledge, and proposes four modes of knowledge conversion of all knowledge in a controlled environment [9][10]. According to the theories of SECI model in conjunction with characteristics of scientific research on remote sensing quantitative retrieval, we adopt the SECI model as follows.

- Socialization: sharing experiential knowledge of the individuals within remote sensing field. Personal merged knowledge system enables individuals to formulate opinion on issues needing improvement, but it could not be expressed specifically at this stage, for instance, an immature idea or speculation.
- Externalization: researchers analyze the characteristics of the existing models, set reasonable assumptions and then improve these from all possible aspects by virtue of their own scientific research capabilities. Eventually the assumptions and ideas transfer into a new model or theory.
- Combination: a stage for model integration and sharing process, which helps researchers evolve a comprehensive understanding of models about associated applications.
- Internalization: the beginning of next round of socialization, externalization, combination and internalization, triggering a spiral development of knowledge transmission and creation.

4. ONTOLOGY BASED KNOWLEDGE REPRESENTATION

Ontology as "*specification of a conceptualization*" [11] is often used to capture domain knowledge explicitly to achieve semantic interoperability. Thus, quantitative remote sensing retrieval models can be driven by the knowledge represented in geospatial and application-specific ontology. We referred to the Ontology for Geospatial Processing Modeling [3] to represent knowledge which include four main ontology as follows: general ontology, domain-specific ontology, geospatial data ontology, geospatial process ontology. General ontology represents the core upper level vocabulary for describing general concepts independent of qualitative remote sensing. Domain-Specific ontology describes the problem space which qualitative remote sensing retrieval models represent. Data ontology provides the scientific meaning of heterogeneous data compliant with the OGC standard of metadata. Process ontology presents the processes' features, classification, internal structure, constraints and relations.

According to RDF (Resource Description Framework), we propose a framework in order to describe the ontological knowledge including invoking restrictions of remote sensing quantitative retrieval models. Unlike a single model network application, users usually are not familiar with the internal mechanism of distributed heterogeneous model resource. In order to integrate these models for a collaborative computing remote sensing quantitative retrieval application, users should be informed of the knowledge on two aspects besides the data processing function of model. One is the logical sequence of the model on each node and the other is whether the former nodes have provided essential parameters before the current one is activated.

Given all the considerations, we descript the model from nine aspects as follows: basic information, function description, data resource, resolution, application category, input parameters, output parameters, constrains of former nodes and evaluation index.

- Basic information – unique identification, URL, developer, version number, copyright and operating environment.
- Function description – elaboration of model function, including theory, assumed condition, prior knowledge, the effect after data processing.
- Data resource – three-layer directory, the main directory divided based on the sensor type, the subdirectory according to data set products and the last directory on basis of band.
- Resolution – specified or universal resolution.
- Application category – applied to specific or common quantitative retrieval application, divided according to the retrieval parameter.
- Input parameters – essential parameters list, including relationship of each with other models.
- Output parameters – output parameters list, including relationship of each with other models.
- Constraints of former nodes – definition of essential processes before the execution.
- Evaluation index – to measure the quality of the model.

Input/output parameters and constraints of former process are the most important elements for the semantic matching, which indicate the relationship between processes. In order to enhance the reuse of models, we divide pre-processing into eight modules in accordance with the generality of various remote sensing quantitative retrieval applications. They are information extraction (from raw data), radiometric correction, atmospheric correction, roi (region of interest), image mosaic (to extract region of interest), geometric correction, cloud mask and resize (unify the resolution of multiple-source data).

5. A CASE OF KNOWLEDGE REPRESENTATION

We build a model-sharing platform named RSWFS (Remote Sensing Workflow System) for remote sensing quantitative retrieval, a java program which enables to integrate models and customize applications. Figure 2 shows the example interface of RSWFWG. The left side bar shows the processes dividing into three categories: data input, data process, data output. Users can drag and drop the processes as their research logic to present a qualitative remote sensing retrieval model. Arrows between the processes represent their successive relationship. Once a model is constructed, users can submit it to the background server supported by a high performance platform. The server receives the request and send back the result to the user after appropriate analysis and transmission Knowledge representation is the core thread throughout the whole design of this platform. All involved processes' detail description, partition granularity and classification are compliant with the ontology for geospatial modeling. Combination and execution of the models include application of knowledge - both explicit knowledge and tacit knowledge.

6. CONCLUSION

In this paper, we have presented a knowledge sharing mode for remote sensing quantitative data retrieval. According to the characteristics of the quantitative remote sensing, we have implemented and extended SECI model. By making a reference to the ontology for geospatial modeling, we abstracted the quantitative remote sensing ontology and summarized nine major aspects of quantitative remote sensing model. We have designed and implemented a platform to verify the knowledge representation effectively.

REFERENCES

[1] L. Di. Customizable, "Virtual Geospatial Products at Web/Grid Service Environment," *Proceedings of IGARSS held in Seoul, South Korea,* 25-29 July 2005.

[2] L. Di, "GeoBrain-A Web Services based Geospatial Knowledge Building System," *In Proceedings of NASA earth science technology conference held in Palo Alto, USA*, 22-14 June 2004.

[3] L. Di, P. Zhao, W. Yang, and P. Yue, "Ontology-driven automatic geospatial-processing modeling based on web-service chaining," *Proceedings of the sixth annual NASA earth science technology conference*. pp.27-29, 2006.

[4] J. Tang, Y. Xue, T Yu, Y. Guan, G. Cai , and Y. Hu, "Aerosol Optical Thickness Determination for Land Surface from MODIS data," *Science in China (Ser. D Earth Sciences)*, vol.49, no.6, pp.641-649, 2006.

[5] Y. Xue, X.W. He, H. Xu, J. Guang, and L.K. Yang, "An aerosol optical depth dataset at 1km resolution over mainland China retrieved from satellite data," *In proceedings of SPIE Asia-Pacific Remote Sensing, held in Kyoto, Japan*,29 October-1 November 2012.

[6] P. Michael, Personal Knowledge: Towards a Post-critical Philosophy, *University of Chicago Press*, 2012.

[7] P. Michael, "The Tacit Dimension," *Garden City, New York*, 1967.

[8] E.A. Smith. "The role of tacit and explicit knowledge in the workplace," *Journal of knowledge Management*, vol. 5, no 4, pp. 311-321, 2001.

[9] K.U. Koskinen, P. Pihlanto, and H. Vanharanta, "Tacit knowledge acquisition and sharing in a project work context," *International Journal of Project Management*, vol.4, no 21, pp. 281-290, 2003.

[10] J.I. Rojas-Sola, M. Castro-Garcia, and M.D.P. Carranza-Canadas, "Contribution of historical Spanish inventions to the knowledge of olive oil industrial heritage," *Journal of Cultural Heritage*, vol.13, no. 3, pp. 285-292, 2012.

[11] T. Gruber, "A translation approach to portable ontology specifications," *Knowledge Acquisition*, vol. 5, Issue 2, pp. 199-220, 1993.

[12] J. Dong, Y. Xue, Z.Q. Chen, J. Liu, Y.J. Li, and T.T. Hou, "A Study of Grid Workflow Dynamic Customization for Remote Sensing Quantitative Retrieval," *In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) , held in Munich, Germany* , 22-27 July 2012.

[13] J. Chen, H. Wu, et al, "Temporal logic and operation relations based knowledge representation for land cover change web services," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 83, pp.140－150, 2013.

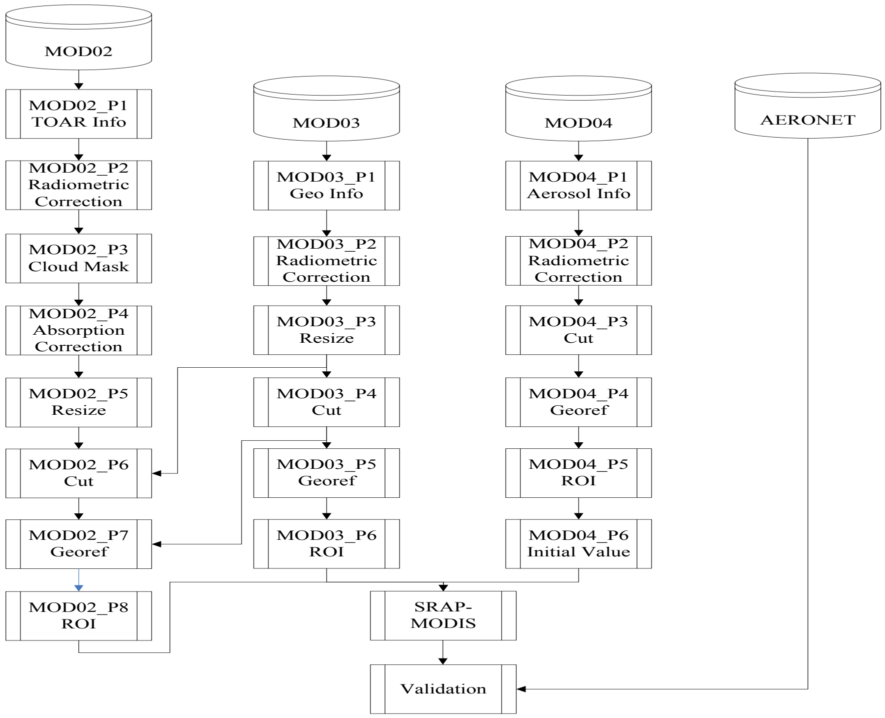

Figure. 1. Diagram of SRAP model

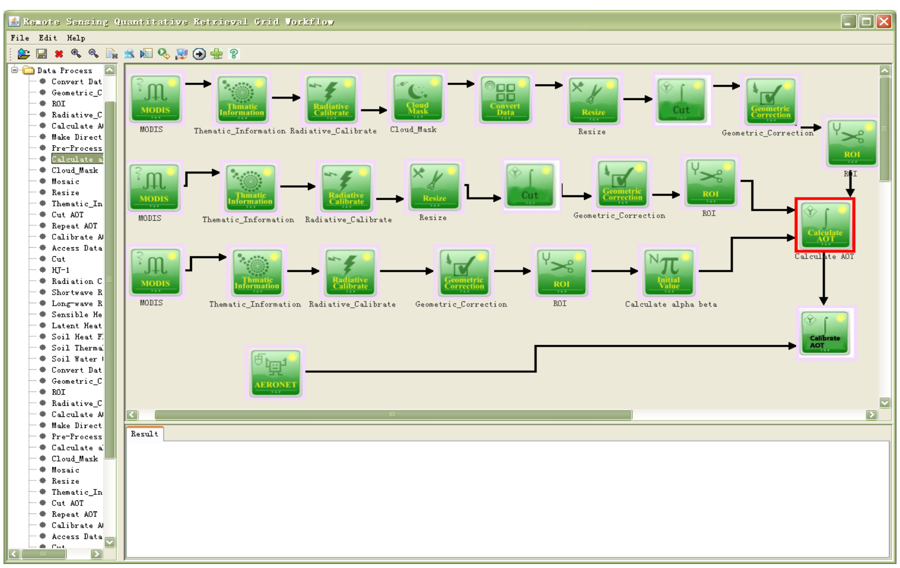

Figure. 2. Interface of the RSWFS

Table1 Former processes of SRAP method

	Information Extraction	Resize	Roi	Geometric Correction	Image Mosaic	Cloud Mask	Radiometric Correction	Atmospheric Correction
MOD02	Y	Y	Y	Y	Y	Y	Y	N
MOD03	Y	Y	Y	Y	Y	N	Y	N
MOD04	Y	Y	Y	Y	Y	N	Y	N