An Ontology Framework for Quality of Geographical Information Services

Richard Onchaga
ITC*
P.O Box 6
7500 AA Enschede, The Netherlands
onchaga@itc.nl

Ing Widya
University of Twente
P.O Box 217
7500 AE Enschede, The Netherlands
widya@ewi.utwente.nl

Javier Morales
ITC
P.O Box 6
7500 AA Enschede, The Netherlands
jmorales@itc.nl

L.J.M Nieuwenhuis
University of Twente
P.O Box 217
7500 AE Enschede, The Netherlands
l.j.m.nieuwenhuis@ewi.utwente.nl

ABSTRACT

In recent years, there has been much research on ontologies for geographic information (GI) services. But to date, focus has been on semantics of data and operations. Much less attention has been given to semantics of quality of GI services. In addressing this gap, this paper proposes an ontology framework for quality of GI services. The framework has an upper ontology, two primary domain ontologies and potentially many application ontologies. The ontologies provide concepts for unambiguously defining and reasoning about quality of GI services. We see the main strength of the framework as being its holistic view of quality; it integrates data quality and quality of service (QoS) which hitherto have been defined separately. Besides, the framework builds on established data quality and QoS concepts which guarantees its relevance and acceptance.

General Terms

Standardization and Interoperability for GIS

Keywords

Quality-aware service chaining, ontology framework, spatial data quality, quality of service

1. INTRODUCTION

In recent years, there has been much research on ontologies for geographic information (GI) services [7, 6]. In a GI services environment, ontologies are required to explicate semantics of geographic data, geoprocessing operations as well as qualities of individual datasets and GI services. To date however, focus has been on semantics of data and operations with little attention being given to quality semantics.

In a marketplace with diverse users and potentially many GI services to address user needs, quality is an important factor of differentiation. Moreover, under such circumstances different levels of quality are possible and at different costs. Ideally therefore, qualities of individual GI services (and datasets) as well as quality requirements of users should be considered during service chaining. We call service chaining in which quality is explicitly considered *quality-aware service chaining*.

Quality-aware service chaining, although desirable, poses new challenges. In particular, the problem of homonyms and synonyms is likely to occur in quality specifications because service users and providers define quality differently and resolving this problem needs a consistent set of shared quality ontologies. Accordingly, this paper proposes an ontology framework for quality of GI services. The framework has an upper ontology, two primary (data quality and QoS) domain ontologies and potentially many application ontologies. Together, the ontologies offer concepts to be used to unambiguously define and reason about quality of GI services. We have implemented the framework and successfully tested resulting ontologies for consistency. We see the main strength of the framework as being its holistic view of quality, i.e., it integrates data quality and QoS which hitherto have been defined separately. Besides, the framework builds on established data quality and QoS concepts which guarantees its relevance and acceptance.

2. QUALITY MODEL FOR GI SERVICES

Traditionally, quality in the GI domain is discussed in terms of intrinsic data quality. Data quality defines fitness-for-use of a dataset in a given context [1]. But in a distributed environment, intrinsic quality will not define the

*International Institute for Geo-Information Science and Earth Observation

ACM GIS November 5-7, 2008. Irvine, CA, USA

utility of a service fully. Aspects of distributed data access and processing also influence the utility information and contribute to the quality of service (QoS) a user perceives. While data quality concerns tangible properties of data e.g. accuracy, completeness and currency, QoS concerns intangible quality properties associated with access to data e.g. performance and security.

2.1 Intrinsic data quality and QoS

Data is said to be *fit-for-use* if it is meets requirements of the target application. Intrinsic data quality is defined along one or more quality dimensions and researchers have identified many data quality dimensions, see e.g. [10]. Quality dimensions for geographic data are a.k.a spatial data quality elements [5] which include completeness, logical consistency, positional accuracy, temporal accuracy and thematic accuracy. Quality of service concerns quality aspects related to access of a GI service. QoS is evaluated along one or more QoS dimensions and researchers have identified several QoS dimensions for web services, see e.g. [8]. We believe these are also relevant to GI services and they include performance, reliability, availability, security and reputation.

2.2 Quality propagation

Intrinsic quality of a dataset is a function of the quality of its source data and the processing operations which transformed source data to create it. Processing operations amplify and transfer errors in data such that source and output data have different error properties. The process of error amplification and transference is called error propagation and is unique for a data quality element and processing operation and is defined in an error propagation model. Given an operation, source data with known error characteristics and corresponding error propagation models, error characteristics of output data can be estimated. Meanwhile, a service chain is a set of GI services executing according to a predefined flow control structure. Individual services in the chain are capable of different levels of QoS and the QoS of the chain is a composition of qualities of individual services in the chain. How QoS is composed is defined in a QoS composition model. QoS composition models are useful for predicting end-to-end QoS capability of a service chain.

3. MOTIVATING EXAMPLE

We motivate the ontology framework with an on-line orthophoto service. The service delivers orthophoto maps on-demand. For brevity, we assume the orthophoto service is realized by a service chain, with each GI service in the chain performing one of a set of activities, i.e., acquire imagery, rectify imagery and overlay imagery with features and annotations. A process model for the orthophoto service is depicted by the UML activity diagram in Figure 1.

In an operational environment, GI services advertise their qualities using domain- or application-specific dimensions. A user will similarly specify quality requirements. Based on the specifications, candidate GI services are selected and chained. For instance, in the example below, the user expects a response time that is no more than 10 minutes, a spatial resolution that is no higher than 10m and no lower than 30m and a total cost no higher than $100. For brevity, we assume that A, Q and P are the set of GI service which satisfy these requirements. In selecting A, Q and P, it first must be resolved that Throughput time, Performance and Speediness refer to the same aspect of quality. Cost, Price/Scene and Price/Use as well as Resolution and Pixel Size must similarly be resolved. Moreover, it should be known how respective qualities aggregate so as to check that candidate services satisfy user requirements. For instance, it is evident that (3+5+1) minutes $\leq$ 10 minutes. Also, $85+$5+$3.50 $\leq$ $100. It is notable that in this case a 30m pixel size image shall be used; using a 10m pixel size image will not fully satisfy user requirements.

$$\begin{aligned}
\texttt{User: Response time} &\leq 10\,\texttt{minutes}; \texttt{Cost} \leq \$\,100;\\
10\,\texttt{m} &\leq \texttt{Resolution} \leq 30\,\texttt{m}\\
\texttt{Imagery Service A: Throughput time} &\leq 3\,\texttt{minutes};\\
\texttt{Price/Scene}\,(30\,\texttt{m Pixel Size}) &= \$\,85;\\
\texttt{Price/Scene}\,(10\,\texttt{m Pixel Size}) &= \$\,125\\
\texttt{Rectification Service Q: Performance} &\leq 5\,\texttt{minutes};\\
\texttt{Price/Use} &= \$\,5\\
\texttt{Mapping Service P: Speediness} &\leq 1\,\texttt{minutes};\\
\texttt{Cost} &= \$\,3.50
\end{aligned}$$

From the example, quality ontologies should help answer the following questions: for a quality dimension, what are its synonyms in domains or applications of interest? what is the quality of a specified GI resource along a given dimension? given a set of quality measures, which measures are compatible? and, given a set of compatible quality measures, how can they be aggregated?

4. ONTOLOGY FRAMEWORK

The ontology framework proposed defines concepts that can help answer the above questions. It has an upper ontology, two primary domain ontologies and possibly several application ontologies in a layered structure. The framework was implemented with ontologies defined using Protégé ontology editor [3] and checked for consistency using RacerPro reasoner [4]. Resulting ontology files can be accessed at [9].

The Upper ontology defines generic concepts for describing quality. Concepts defined in the upper ontology are refined for use in domains and applications as necessary. The main concepts of the upper ontology are quality measure, quality dimension, unit of measure (for a dimension), direction (of a dimension) and domain (of a dimension) and user requirement. Figure 2 illustrates the upper ontology. Domain ontologies define concepts used in specific domains by specializing concepts of the upper ontology. The domains of interest to this study are data quality and QoS with corresponding ontologies as data quality ontology and QoS ontology respectively. The data quality ontology defines concepts to be used to define quality of geographic data. The main concepts of the ontology are data quality measure, data quality element, error propagation model and computational model quality. On its part, QoS ontology defines concepts that are used to describe QoS. The core concepts of ontology are QoS dimension, QoS measure and QoS composition model.

5. RELATED WORK

In GI handling, data quality is a standardized subject [5]. Data quality is defined in terms of its intrinsic properties without considering aspects that influence its access. More

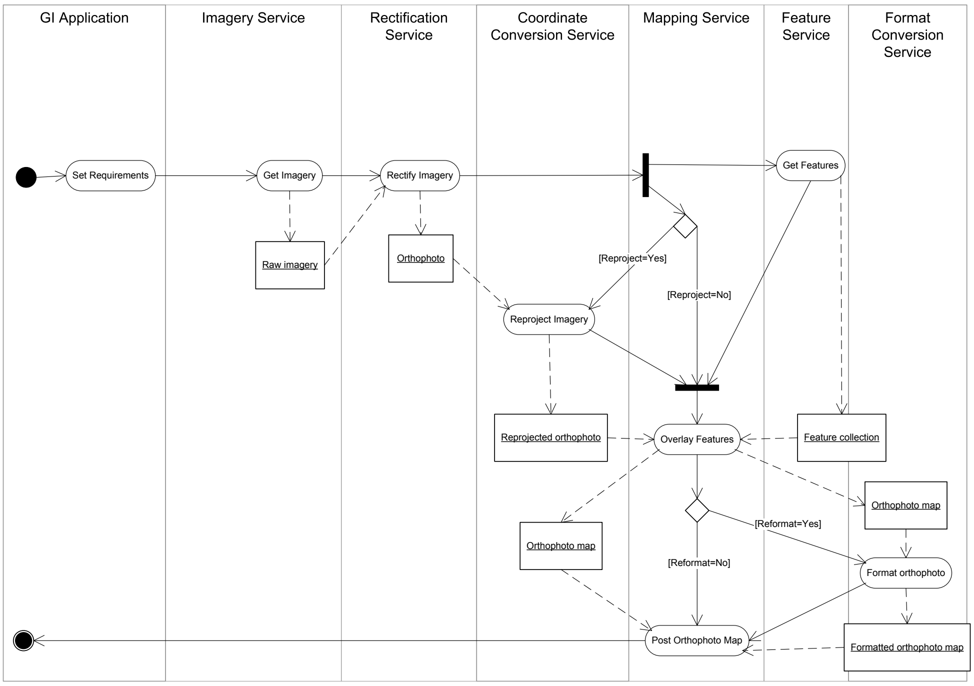

Figure 1: Orthophoto service process model

recently, studies on performance have been reported, see e.g. [11] but these have centered on performance and ignored other QoS properties. This paper extends the definition of quality beyond intrinsic data quality and performance to include QoS in a broader sense.

In recent years, several research studies on formal ontologies for GI services have been reported e.g. [7, 6]. The studies mainly address semantics of data and operations but do not address quality. The framework we propose augments data and operation ontologies towards enabling truly customized GI services.

QoS is well documented with many QoS ontologies defined, see e.g. [2]. For flexibility, authors adopt a modular and layered structure for the ontologies– a structure also used in this paper. While similar in many respects to common QoS ontologies, the framework we propose differs in two important respects. First, it defines generic concepts and relations in the upper ontology which can used to define quality in any domain or application. Further, it offers concepts for describing data quality which are not covered by contemporary QoS ontologies.

6. CONCLUSION

In this paper we proposed an ontology framework for quality of GI services. The framework offers a consistent set of ontologies that can be used to semantically enrich quality specifications and enable quality-aware service chaining.

7. REFERENCES

[1] A. Agumya and G. J. Hunter. Determining fitness for use of geographic information. *ITC Journal*, 2, 1997.

[2] G. Dobson, R. Lock, and I. Sommerville. Qosont: a qos ontology for service-centric systems. *Software Engineering and Advanced Applications, 2005. 31st EUROMICRO Conference on*, pages 80–87, Aug.-3 Sept. 2005.

[3] S. C. for Biomedical Informatics Research. Protégé ontology editor and knowledge acquisition system. http://protege.stanford.edu.

[4] R. S. GmbH. Racerpro version 1.9. http://www.racer-systems.com.

[5] ISO/TC211. *ISO 19113 Geographic Information - Quality Principles*. ISO, 2002.

[6] R. Lemmens, A. Wytzisk, R. de By, C. Grannel, M. Gould, and P. van Oosterom. Integrating semantic and syntactic descriptions to chain geographic services. *IEEE Internet Computing*, pages 42–51, September|October 2006.

[7] M. Lutz. Ontology-based descriptions for semantic discovery and composition of geoprocessing services. *Geoinformatica*, 11(1):1–36, March 2007.

[8] D. A. Menascé. QoS issues in web services. *IEEE Internet Computing*, pages 72–75, November|December 2002.

[9] R. Onchaga. Quality ontologies for geographic information services. http://www.geocities.com/onchagamoses/quality-ontology/, 2008.

[10] Y. Wand and R. Y. Wang. Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, 39(11):86–95, 1996.

[11] C. P. Yang, D. W. Wong, R. Yang, M. Kafatos, and Q. Li. Performance-improving techniques in web-based GIS. *Int. J. Geographical Information Science*, 19(3):319–342, 2005.

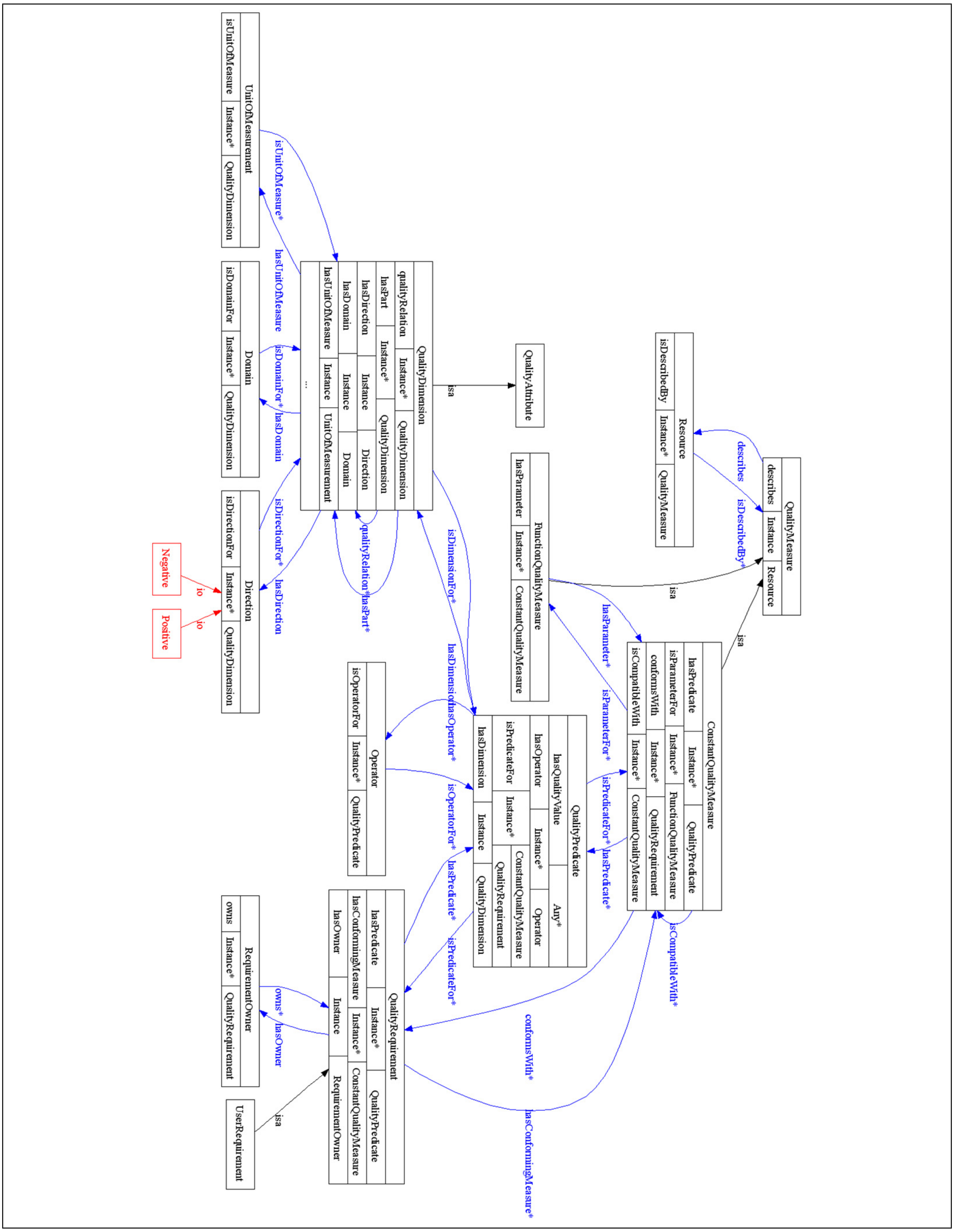

Figure 2: A schematic representation of the Upper quality ontology