

Geospatial Data Qualities as Web Services Performance Metrics

Ganesh Subbiah Ashrafal Alam Latifur Khan Bhavani Thuraisingham
Department of Computer Science, Erik Johnson School of Engineering and Computer Science
University of Texas at Dallas

ganesh.subbiah@student.utdallas.edu, malam@utdallas.edu,
lkhan@utdallas.edu, bhavani.thuraisingham@utdallas.edu

ABSTRACT

Service discovery is the crucial phase in the emerging Geospatial Semantic Web to select functionally similar services for the user query. Quality of Service (QoS) based service discovery, popularly studied in traditional Web Services, applies also to Geospatial Web Services. QoS allows service clients to fine-tune their search according to their specific needs and criteria. In high-performance service-based geospatial applications, it becomes an interesting research challenge to identify geospatial parameters to further improve the search process. In this paper we have proposed a set of geospatial criteria that can be used alongside the regular QoS parameters in service discovery and invocation. We show that using this novel approach of incorporating domain-specific drill-down information in addition to the commonly used QoS parameters yield more accurate and trustable Web services platform. We use the proposed geospatial parameters as performance metrics in the experimental evaluation of our application. The parameters reflect geospatial data quality attributes already standardized and well-studied in geospatial literature.

Categories and Subject Descriptors

D.2.8 [Software Engineering]: Metrics – *complexity measures, performance measures, process metrics.*

General Terms

Algorithms, Performance, Design, Reliability, Experimentation, Security.

Keywords

Geospatial Semantic Web, Semantic Web Services, Security, Trust, SOA, OWL-S, RDF

1. INTRODUCTION

Web services are increasingly seen as an invaluable part of any large-scale data query and dissemination strategy. The rise of Service Oriented Architecture (SOA) to provide intra- and inter-

domain business services has ensured the rapid growth of Web services as the primary delivery platform. Business can query, find, and invoke specific services to perform their tasks instead of relying on bulky applications with superfluous features. Web services are a perfect suit for the geospatial domains since geospatial features are easy to modularize and serve to clients. As a result, clients can retrieve only the pertinent data according to their need.

Geospatial web services have been an active area of research in the context of geospatial non-interoperability problems. The collaborative effort by the industry and federal geospatial clearinghouses has focused on the standardization process to mitigate the non-interoperability problems. Although the importance of geospatial web services is well established, their efficiency is often questionable. Geospatial data tends to be voluminous even for few features; consequently on-the-fly data fetching becomes infeasible. Moreover, the data comes in various modalities even though they represent the same base facts. For instance, aerial imagery can be viewed at different resolution and vector data can be represented in different granularities. Then there is the issue of data quality that further exacerbates the efficiency of geospatial web services. To eliminate the above impediments, web services are incorporated with the Quality of Service (QoS) parameters that provide a baseline contract of what a client wants and what to expect from a service provider.

The issue of QoS has provided a major area of Web services research ([3],[4]). In [3][4], QoS based service selection is used to find trustworthiness of web services. The common theme in the geospatial QoS literature is to use the regular QoS parameters to efficiently exchange geospatial data [5]. The addition of domain information in the QoS values has been overlooked by researchers so far. Also there is not much work done on using geospatial specific QoS for estimating the trustworthiness of the geospatial web service

Our experience in building end-to-end geospatial web services frameworks [1,2], we have found that the client requirements revolve around four major threshold types: completeness, resolution, accuracy, and data type [6]. While there are other requirements as well, these four appear on a consistent basis. The completeness, resolution, and accuracy criteria pertain to qualitative side of geospatial data, whereas data type refers to the format of the data. Our approach is to combine these four criteria alongside the generic QoS parameters to yield a more customizable and client-centric geospatial web services platform. We refer to these four criteria as GQoS- Geospatial Quality of Service metrics.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

ACM GIS '07, November 7–9, 2007, Seattle, WA, USA.
Copyright 2007 ACM 978-1-59593-914-2/07/11...\$5.00

In this paper, we propose a framework which provides a mechanism to select trustworthy geospatial web services based on geospatial quality parameters. The application is based on our work on semantically annotated geospatial web services discovery. We develop an application called DAGIS (Discovery of Annotated Geospatial Information Services)[2], which we augment with GQoS and perform experimental evaluations to show its usefulness in identifying trust measures dynamically and to eliminate untrustworthy services for the query. This DAGIS framework provides a methodology to realize the semantic interoperability both at the geospatial data encoding level and also for the service framework. DAGIS is an integrated platform that provides the mechanism and architecture for building geospatial data exchange interfaces using the OWL-S Service ontology. Coupled with the geospatial domain specific ontology for automatic discovery, dynamic composition and invocation of services, DAGIS is a one-stop platform to fetch and integrate geospatial data.

The rest of the paper is organized as follows. Section 2 presents the DAGIS architecture. The proposed GQoS metrics are described in section 3. Section 4 describes service selection algorithm using GQoS parameters. The experiments are reported in section 5.

2. DAGIS PLATFORM for GEOSPATIAL SERVICES

In DAGIS ([1]), we focus on devising an improved query mechanism through semantic annotations. The application allows clients to query on a visual interface for geospatial data. The returned results can be intermingled with other types of data if requested. The results retrieved by a client can be displayed on the interface or stored on disk files. This section describes the DAGIS architecture.

2.1 Motivating Scenario

“Find movie theaters within 30 miles of 75080” is a query posed by users on current geospatial information systems and search engines. This query is an example of the type of requests carried out by service providers on the web. Service providers would often embed or layer the geospatial data in other kinds of data (e.g., medical, temporal, transactions etc.). The following sections describe how DAGIS platform handles queries of this nature.

2.2 Service Selection and Discovery

First, a query profile is generated based on the client request. The profile contains the functional and QoS metrics of the specified parameters in the client request. These requirements are used by the Matchmaker agent for selecting the appropriate service providers. In this phase, a DAGIS application module, henceforth referred to as DAGIS agent or simply agent, communicates with the Matchmaker agent for geospatial service selection (Figure 1). Prior to the service discovery, the agents of each service provider advertise the respective OWL-S service profiles to the Matchmaker. The Matchmaker in our framework does capability based reasoning using the Pellet OWL-DL reasoner. The implemented Matchmaker for this framework is based on the OWL-S MX Matchmaker, a hybrid Matchmaker that complements logic based reasoning with approximate matching based on syntactic IR based computations.

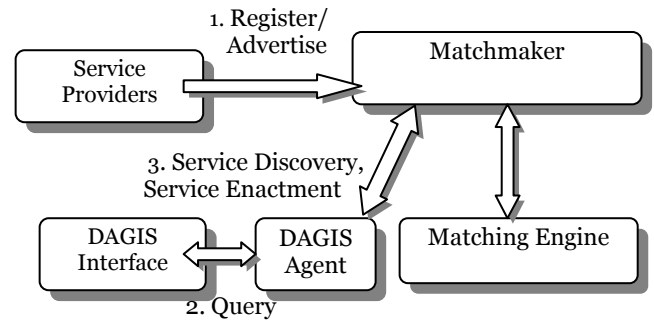


Figure 1. DAGIS System Architecture

2.3 Service Invocation

In this phase, the DAGIS agent has the selected Service Provider’s Uniform Resource Identifier (URI) from the discovery process and invokes the provider by calling one or more business methods on the URI. The service provider agent uses the same domain ontology as the DAGIS agent for semantic annotations of its services. The DAGIS agent does the invocation of the service through OWL-S grounding. The OWL-S grounding in turn uses WSDL grounding to invoke the Web Service using AXIS in our framework.

3. GEOSPATIAL QUALITY of SERVICE (GQoS)

We have proposed a set of four geospatial attributes, commonly used to specify data quality for various standards, to incorporate into our base framework (i.e., DAGIS). They augment the generic QoS parameters to allow geospatial users more precise control over their query. There are many advantages in using this approach. Traditional Web services provide the modularity but take away the ability to precisely control the use of the data. To get around this problem, one can retrieve a large amount of data from a service provider and perform offline filtering or various types of modifications themselves. However, this is a very inefficient and time-consuming procedure since a lot of processing is done post hoc. The GQoS parameters allow clients to restrict the types of service providers it is interested in before any processing on the data is done. If there is no provider available that matches the client criteria, then the client can alter the query and resubmit. These GQoS Parameters are added as OWL-DL classes to our QoS Ontology described in our previous paper [1].

In this section, we describe the following GQoS parameters: Accuracy, Resolution, Completeness, and Types.

3.1 Accuracy

Accuracy of geospatial data is defined in terms of (Attribute, Value) tuple, where attribute refers to a geographic concept/object and the Value is its measurement. We assume geospatial service providers provide data that conform to such tuples. We also assume that there is an objective assessment of all concept values. Governmental agencies, for example, would be assumed to have

the most accurate object values in the event that there are multiple values for a geographic object.

3.2 Resolution

Resolution refers to the amount of detail that can be determined in space, time or theme. Both image and vector data have resolution properties. Image resolution generally refers to pixel details where more pixels per unit of an image mean better clarity. Vector data can be represented in either fine or coarse granularity. The coarser the data is, the less information is available about vector points of an object's shape. Resolution is also related to accuracy because the level of resolution affects the database specification against which accuracy is assessed.

3.3 Completeness

Completeness refers to the absence of omissions in a provider database. Completeness is distinct from accuracy in that the errors that result in lack of completeness are not incorrect encoding of object values. Instead, when a service provider fails to keep its database updated with latest data is considered to have incomplete data. For instance, the road Atlas of 2006 contains data about roads and highways built since the previous Atlas editions were published. As a result, the 2006 version would contain more complete information than the one from year 2000.

3.4 Data Types

Data types refer to the format of desired data. Even though the area of geospatial data interoperability has made a lot of progress, various reasons still exist that lead clients to request specific type of data format. For instance, although Open Geospatial Consortium (OGC) has been pushing Geography Markup Language (GML) as a standardized data exchange platform, not all geospatial applications support it. As a result, it would be rather inconvenient for a user of such an application to request data from a provider only to end up with GML data. If the user could specify that along with other requirements in the query, he can avoid spending time on retrieving useless data.

4. PROPOSED ALGORITHM to IMPLEMENT GQoS PARAMETERS

During the Semantic Service Discovery Phase[2.2], the query profile of the user is submitted to the matchmaker for determining the functional matches from the set of published services. The Matchmaker returns a set of functionally similar services if the query to be solved involves single service provider; otherwise returns a dynamically composed service if the query requires service orchestration.

4.1 New Service Discovery Algorithm

To incorporate the GQoS values, we add a step to the DAGIS service discovery algorithm. The new algorithm operates as follows.

1. Service providers publish profiles to Matchmaker
2. Generate query profile
3. Find semantically similar services for the query using the functional parameters :input and output parameters
4. If there is no such service from step 3, dynamically compose complex service using the services registered using DAGIS Composer Algorithm [2]

5. Sort the Functionally Similar Semantic Services using the GQoS Algorithm (see Figure 3)
6. Return the URI of the best Service from step 5 to user

We will describe the approach developed by us for performing the Step 5 of the service discovery algorithm. The QoS selection differs when we have a dynamic composition that involves computing the aggregate QoS values of the services dynamically, which is also one of our contribution in this paper.

4.2 GQoS Algorithm

Interaction Model: The Environment is comprised of registered service providers $S_1, S_2 \dots S_j$, Users $U_1, U_2 \dots U_i$, matchmakers $M_1, M_2 \dots M_k$. In our interaction model we assume only one matchmaker. We employ special monitoring services which get the user reports on QoS relevance feedback which are called Trust Monitors $TM_1, TM_2 \dots TM_i$. Matchmaker can also additionally act as Trust Monitor,

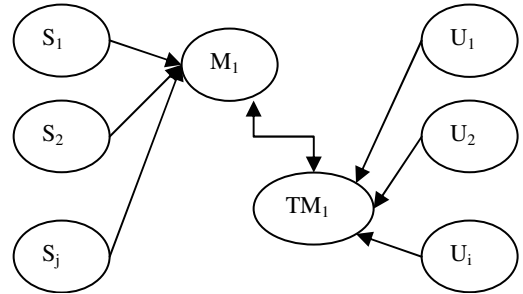


Figure 2. Interaction Model

```

User Query List UQ = {(uq1, r1), (uq2, r2)
... (uqn, rn)}
TargetMatch // Number of concept matches
required
Gval = 0 for all services
1.  $\forall S_j$  in Functional Match Set F
2. dist = 0.0
3.  $\forall q_i: q_i = \text{quality concept in } uq$ 
4. If  $q_i$  matches with a concept in  $sq_j$ 
5. conceptmatch = conceptmatch + 1
6. dist +=  $|r_i - p_i|$ 
7. If concept match  $\geq$  TargetMatch then
8. Gval = diff/conceptmatch
9. Return F sorted by ascending order of
Gval scores.
  
```

Figure 3: GQoS Similarity Match Algorithm

Service providers publish their QoS values $(sq_1, p_1), (sq_2, p_2), \dots$ where (sq_i, p_i) are vector pairs of concepts and their values. Users provide the QoS requirements for every query as $(uq_1, r_1), (uq_2, r_2), \dots$ where (uq_i, r_i) are vector pairs of concepts and user required values. GQoS vector values p_i, r_i are fuzzy values which are in the range $[1,5]$. 1 is the worst GQoS support available and 5 is the best support available for that GQoS parameter.

In the First Phase, for each registered Service Provider j in the functional match set F of the Query Q , a G_{val} is evaluated using the advertised QoS parameters. G_{val} is the Manhattan distance averaged over the number of quality concept matches between the user requirement and the service provider advertised GQoS values.

$$G_{val} = \sum (r - p) / \text{conceptmatch}$$

The GQoS Similarity Match Algorithm is illustrated in figure 3 to select a set of services. All the Service providers are set with $G_{val} = 0$ and the target concept matches between query and service provider concept is set to a constant (is 3 in our experiments). In Step 1 for every service S_j returned from Functional Set F returned from Matchmaker. The similarity between r and p vectors is measured using Manhattan Distance. For every quality concept q_i in Vector u_q , if there is a ConceptMatch (exact, subsumes) with a concept in sq_j , conceptmatch is incremented. The diff is updated for this match, In step 7 we check if there are at least target number of matches for meeting the user requirement, we compute the G_{val} as average distance over the concept matches in step 8. Step 9 returns the F in ascending order of G_{val} .

In the second phase of the GQoS measurements, we use the user feedback to update the advertised GQoS parameters of the selected service S_j as follows. All the user reports pertaining to the similar query Q posed is aggregated here in this phase. The user feedback list UF of every user is evaluated as shown in Figure 4.

In our model, user reports are considered to be credible as only authenticated users of the system can log on to the system for service discovery. The evaluation of the credibility of user values reported is not in scope of our work. We assume that the Service Providers who publish their service descriptions to the

```

Aggregate Feedback Vector FV;
For every Service Provider Sj
1. Read every User Feedback List UiF received =
{(uq1, f1), (uq2, f2) ... (uqn, fn)} where i=1:n
2. FV = FV + {(uq1, f1), (uq2, f2) ... (uqn, fn)}
3. End For
4. FVavg = FV / n
5. Update each QoS parameter sqj of Sj as
    pj = pj (1 - Fv) + Pj

```

Figure 4: GQoS Propagation Algorithm

matchmaker do not cancel their registration during the interaction for at least a certain number of iterations (say 10) to facilitate the catching of untrusted providers. In future, we would maintain logs of the interactions to capture these cancellation scenarios also.

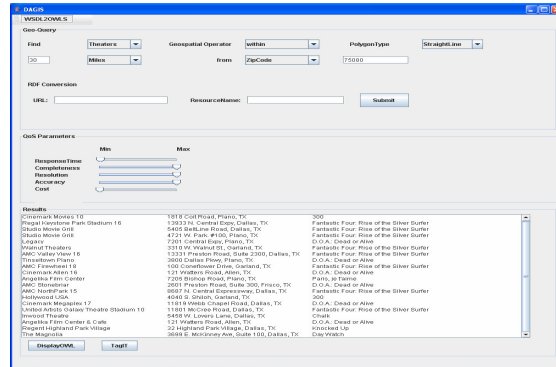
5. EXPERIMENTAL EVALUATION

The experiment and evaluation results are to be shown during the demo at the poster session.

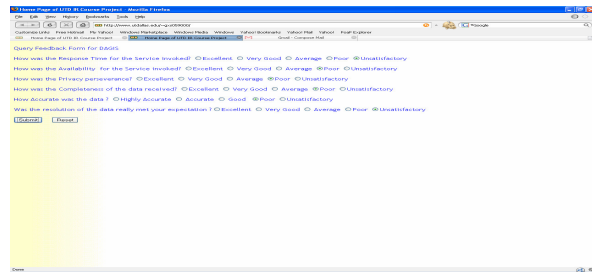
6. CONCLUSION

In this paper, we have successfully proposed geospatial data parameters which are used in the automatic service discovery for emerging semantic enabled geospatial web. The framework proposed and implemented helps to distinguish the untrustworthy service providers by penalizing them using the performance

metrics evaluated by keeping the user in the loop. We are working on further experiments which show the increase in the precision and relevance measures due to these proposed geospatial quality metrics. This work provides an intuitive way to select trustworthy semantic web services using the geospatial data quality parameters along with QoS measures which is novel step towards the building geospatial web of trust.



Appendix A DAgIS Semantic Query Interface



Appendix B User Feedback Form for a Query

7. REFERENCES

- [1] Ashrafal Alam, Ganesh Subbiah and Bhavani Thuraisingham: Reasoning with Semantics-aware Access Control Policies for Geospatial Web Services. 2006 ACM Workshop on Secure Web Services (SWS).
- [2] Ashrafal Alam, Ganesh Subbiah, Latifur Khan, and Bhavani Thuraisingham: DAGIS: A Geospatial Semantic Web Services Discovery and Selection Framework, Second International Conference on geospatial semantics GeoS2007
- [3] Maximilien, E. M. and Singh, M. P. 2004. Toward autonomic web services trust and selection. In Proceedings of the 2nd international Conference on Service Oriented Computing ,ICSOC '04. ACM Press.
- [4] Le-Hung Vu, Manfred Hauswirth, and Karl Aberer: QoS-based Service Selection and Ranking with Trust and Reputation Management, 2005 International Conference on Cooperative Information Systems (CoopIS), Nov 2005
- [5] Wen-jun Li, Shu-neng Zhao, Heng Sun, Xiao-bin Zhang, "Ontology-Based QoS Driven GIS Grid Service Discovery," *skg*, p. 49, Second International Conference on Semantics, Knowledge, and Grid (SKG'06), 2006
- [6] Veregin H and Hargitai P 1995 An evaluation matrix for geographical data quality. In Guptill S C and Morrison J L (eds) Elements of spatial data quality. Oxford: Elsevier 167-188.