Computers & Geosciences 92 (2016) 1–8

Contents lists available at ScienceDirect

Computers & Geosciences

journal homepage: www.elsevier.com/locate/cageo

Case study

Investigating metrics of geospatial web services: The case of a CEOS federated catalog service for earth observation data

Weiguo Han [a,*], Liping Di [b], Genong Yu [b], Yuanzheng Shao [c], Lingjun Kang [b]

[a] Visiting Scientist Programs, University Corporation for Atmospheric Research, 3090 Center Green Drive, Boulder, CO 80301, USA
[b] Center for Spatial Information Science and Systems, George Mason University, 4400 University Drive MS 6E1, Fairfax, VA 22030, USA
[c] Collaborative Innovation Center of Geospatial Technology, Wuhan University, 129 Luoyu Road, Wuhan, Hubei 430079, China

ARTICLE INFO

Article history:
Received 25 November 2014
Received in revised form
17 February 2016
Accepted 8 April 2016
Available online 9 April 2016

Keywords:
Federated catalog service
Catalog service for the web
Geospatial web services
Web service metrics
Earth observation

ABSTRACT

Geospatial Web Services (GWS) make geospatial information and computing resources discoverable and accessible over the Web. Among them, Open Geospatial Consortium (OGC) standards-compliant data, catalog and processing services are most popular, and have been widely adopted and leveraged in geospatial research and applications. The GWS metrics, such as visit count, average processing time, and user distribution, are important to evaluate their overall performance and impacts. However, these metrics, especially of federated catalog service, have not been systematically evaluated and reported to relevant stakeholders from the point of view of service providers. Taking an integrated catalog service for earth observation data as an example, this paper describes metrics information retrieval, organization, and representation of a catalog service federation. An extensible and efficient log file analyzer is implemented to retrieve a variety of service metrics from the log file and store analysis results in an easily programmable format. An Ajax powered Web portal is built to provide stakeholders, sponsors, developers, partners, and other types of users with specific and relevant insights into metrics information in an interactive and informative form. The deployed system has provided useful information for periodical reports, service delivery, and decision support. The proposed measurement strategy and analytics framework can be a guidance to help GWS providers evaluate their services.

1. Introduction

Open Geospatial Consortium (OGC) has made a series of specifications for geospatial Web services (GWS) covering discovery, access, portrayal, and processing of geospatial data. The specifications, like Web Map Service (WMS), Web Coverage Service (WCS), Web Feature Service (WFS), Web Processing Service (WPS), and Catalog Service for the Web (CSW), have been extensively adopted and implemented in the industry and academic community. The OGC standard-compliant services contributed to make geospatial data and computing resources discoverable and accessible over the Web, and greatly facilitated sharing and interoperability of geospatial information from distributed sources. The wide adoption of OGC services raised the need to monitor and evaluate performance of GWS from both service providers and consumers.

As an international interagency organization, Committee on Earth Observation Satellite (CEOS) coordinates satellite Earth Observation (EO) programs between space agencies of its member countries. The goal of CEOS Working Group on Information Systems and Services (WGISS) is to provide EO data management systems and services to worldwide users. Initiated by National Aeronautics and Space Administration (NASA) and National Oceanic and Atmospheric Administration (NOAA) in 2010, CEOS WGISS Integrated Catalog (CWIC) project aims to build a federated catalog system which provides inventory level search results from EO data catalog systems of CEOS members through a standard unified interface (Enloe and Yapur, 2011). Currently, the operational CWIC service (http://cwic.wgiss.ceos.org) has integrated EO data catalog systems from NASA Earth Observing System (EOS) Clearing House (ECHO), NOAA Group for High Resolution Sea Surface Temperature (GHRSST), United States Geological Survey (USGS) Land Surface Imaging (LSI), Brazil National Institute for Space Research (Instituto Nacional de Pesquisas Espaciais, INPE), and Canada Centre for Mapping and Earth Observation (CCMEO). It has been served as one of data sources for the Societal Benefit Area (SBA) of Agriculture and Disasters activities in the Global Earth Observation System of Systems (GEOSS) Architecture Implementation Pilot (AIP) projects (Percivall et al., 2013).

* Corresponding author.
E-mail address: hanwg.bj@gmail.com (W. Han).

http://dx.doi.org/10.1016/j.cageo.2016.04.005

CWIC project stakeholders, sponsors, developers, and partners are most concerned with the questions about metrics of this integrated catalog service, such as "*How many users have accessed CWIC service in the past month?*", "*How long does it take CWIC service to handle GetRecordById request?*", "*Which data collections or datasets of USGS LSI are requested most often?*", "*Which countries are CWIC users from?*", and so on. Such metrics information should be collected, organized, and presented for evaluation of overall performance and impacts of CWIC service. This paper will address these and other related requirements. To make the process smoother and easier, a log file is necessary to track request processing for metrics information collection, a log file parser is needed to retrieve metrics information from the log file and organize them in an easy-to-display format, and a web dashboard is required to represent metrics information in informative tables and charts. And this mechanism can also be extended and applied to other similar Web services.

The reminder of this paper is organized as the following. Section 2 reviews the progresses in evaluation of GWS metrics and introduces the system requirements. Section 3 presents technical approaches in the implementation of CWIC metrics information collection, extraction, and representation. In Section 4, system functions are demonstrated to provide detailed and actionable insights into CWIC metrics. Section 5 discusses the experiences from metrics monitoring and analysis of CWIC. Finally, Section 6 summarizes conclusions and directions for future work.

2. Related works

2.1. Geospatial web services

Many organizations have followed OGC standards to publish their geospatial data, information, and services in an open and interoperable way. These standards are broadly grouped into three categories, i.e. data, processing, and catalog.

Geospatial data services support geospatial data customization and retrieval according to input parameters. For example, WCS services provide access to geographical coverages through standard operations (Whiteside and Evans, 2008), WFS services offer vector data manipulation and retrieval (Vretanos, 2002), and WMS services handle geospatial data rendering and portrayal (de la Beaujardière, 2006). These data services are well supported in both commercial and open source GIS software. The issues related to quality of this kind of service have been addressed by many researchers. From the perspective of service consumers, Zhang et al. (2010) used metrics of precision and recall to evaluate WFS query results; Horák et al. (2011) measured response time, error occurrence, availability, and performance of WMS services by repeating same requests; Wu et al. (2011) presented a new approach to monitor and assess quality of WMS services and developed a mechanism to choose better map layers for decision making support; Gui et al. (2013) leveraged Geospatial Cyber-infrastructure (GCI) components to build a search engine framework for geospatial resources discovery and registry, and developed a quality monitoring and evaluation module to assess accessibility and performance of registered OGC data services. In addition, Giuliani et al. (2013) proposed a new approach to evaluate performance of WFS and WCS services on the server side and provided service providers with guidance on service quality improvement.

Geospatial processing services offer operations for geospatial data transformation and processing derived from geospatial models, algorithms, and applications. The WPS specification defines standard interfaces for discovery of, publishing of, and binding to geospatial process (Schut, 2007), so WPS services can be composed in scientific workflows to perform complex tasks over distributed geospatial resources (Cepicky and Becchi, 2007; Kiehle et al., 2007). Measurements of data transfer fluency and processing control in the workflow can be used to evaluate quality and performance of interoperability (Gorgan et al., 2012). Scholten et al. (2006) analyzed four performance-related factors for geoprocessing services, including caching, network adaptation, data granularity, and communication mode. Sun et al. (2012) developed a prototype system called GeoPWTManager to chain geo-processing services and monitor and visualize performance of these services.

Geospatial catalog services provide geospatial information registry, description, discovery, and access. OGC CSW specification defines standard interfaces to register, publish and search geospatial data, information, and services in the metadata catalogs (Voges and Senkler, 2005). The general query criteria contain spatial extent, temporal range, and dataset identifier, etc. This specification has been adopted and implemented in many applications, like GeoBrain Catalog Federation service (Bai et al., 2007), GEOSS Component and Service Registry (Bai et al., 2012), Group on Earth Observations (GEO) Discovery and Access Broker (DAB) (Nativi et al., 2013), *GeoNetwork* (http://geonetwork-opensource.org), and *deegree* (http://www.deegree.org). However, in comparison with the other two categories of geospatial service, less effort has been devoted to measure quality of catalog services which is critical for both service providers and end users, and there have been fewer publications on this topic so far.

2.2. Web analytics

Web analytics tools collect and display metrics of a website or Web application, and give powerful indicators on its performance, capacity, and availability. In these tools, analytical statistics are performed on the measurements of each aspect of a website or web application to provide information like visits, ranking, and processing time on traffic history. The metrics results are presented in a detailed web traffic dashboard with interactive tables and colorful graphs for decision making support.

Server log file analysis and page tagging are two common technical solutions on visit information collection. In the former method, the log file or database collecting web activities is parsed and analyzed through self-hosted web analytics software. AWStats (http://awstats.sourceforge.net) is an open source web analytics application for processing visit information from the server log file and presenting them visually within static HTML reports. Other open source alternatives to AWStats are Analog, Webalizer, and W3Perl. Google Analytics (http://www.google.com/analytics), belonging to the latter approach, is one of the most popular web analytics programs in the world today. Users only need to embed a snippet of JavaScript tracking code in their web pages, Google Analytics will help them track visitors along with their activities from browser cookies and learn full pictures of their websites, such as where visitors are from and where web traffic comes from. Other free web analytics services from different vendors include Yahoo! Web Analytics (http://web.analytics.yahoo.com/), Bing Webmaster (http://www.bing.com/toolbox/webmaster), Quantcast Measure (https://www.quantcast.com), etc.

Miller et al. (2002) analyzed service metrics using a Quality of Service (QoS) model with three dimensions (i.e. time, cost, and quality). Giuliani et al. (2013) pointed out that quality of downloading services like geospatial data services should be evaluated using three criteria: 1) performance, 2) capacity, and 3) availability, and these criteria can be measured on either server side or client side. Khaled et al. (2010) proposed to enhance metadata information on quality using ISO 19119 standard, and suggested that the GWS quality should be evaluated based on spatial data