



A novel digital information service for federating distributed digital entities



Ahmet F. Mustacoglu ^{a,*}, Geoffrey C. Fox ^{b,c}

^a TUBITAK BILGEM – The National Research Institute of Electronics and Cryptology (UEKAE), Turkey

^b School of Informatics and Computing, Indiana University, Bloomington, IN, USA

^c Community Grids Lab, Indiana University, Bloomington, IN, USA

ARTICLE INFO

Article history:

Received 30 June 2015

Received in revised form

27 July 2015

Accepted 28 July 2015

Available online 8 August 2015

Keywords:

Web services and Service-Oriented computing

SOA

Web 2.0

Data management

Information retrieval and management

Federation and unifications

ABSTRACT

We investigate the performance and the scalability metrics of a Digital Information Service framework that is used for unifying and federating online digital entities by retrieving and managing information located on the web. The Digital Information Service consists of tools and web services for supporting Cyberinfrastructure based scientific research. This system supports a number of existing online Web 2.0 research tools (social bookmarking, academic search, scientific databases, journal and conference content management systems) and aims to develop added-value community building tools that leverage the management and federation of digital entities and their metadata obtained from multiple services. We introduce a prototype implementation and present its evaluation. As the results indicate, the proposed system achieves federation and unification of digital entities coming from different sources with negligible processing overheads.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Information is spread all over the Web in various locations including centralized repositories, web servers and user desktops. Centralized repositories represent the old fashion techniques for resource sharing, whereas completely decentralized systems such as P2P systems allow users to share information without depending on a third party repository. The necessities to find and share information led to development of emergent Web 2.0 applications. These new Web 2.0 applications such as social bookmarking tools introduce a new way of sharing information with respect to the old fashion and P2P systems do. Social bookmarking tools address the challenging problems of finding and sharing

information among small groups, teams and communities. Various types of social bookmarking tools developed their own systems to support different kinds of resources. Flickr, for example, allows the tagging and sharing of photos, del.icio.us the tagging and sharing of bookmarks, BibSonomy, CiteULike and Connotea the tagging and sharing of scholarly publications, YouTube the tagging and sharing of video, and 43Things the tagging and sharing of goals in private life. Some of these tools may not survive in the future, for example, Connotea ended operation in March 12, 2013 due to the growing problems with spam and associated service outages. Note that this type of cases can be thought of as an advantage since our proposed system stores data coming from external services locally and the stored data can be retrieved and exported into supported file formats (.txt etc.). Furthermore, in the case of these tools are not operating continuously (e.g., system down status of few days) our system would continue to operate perfectly. The proposed system stores the data in its own database hence the data

* Corresponding author.

E-mail addresses: afatih.mustacoglu@tubitak.gov.tr (A.F. Mustacoglu), gcf@indiana.edu (G.C. Fox).

Table 1
Status of Web-based popular academic/non-academic services.

Tool name	Academic (A)/non-academic (NA)	Start–end date	Users/data as of now (valid as of 02.01.2015)	Current status
Citeulike	A	11.2004–current	Around 7.4 million articles	Active
Connotea	A	12.2004–03.2013	No information	Retired
BibSonomy	A	Early 2006	Around 3.4 million links	Active
Google Scholar	A	11.2004–current	All scholarly literature	Active
Microsoft Academic Search	A	11.2009–current	Over 4.1 million papers	Not updated
ResearchGate	A	05.2008–current	Over 3 million members	Active
IEEE Xplore	A	02.2000–current	Over 3 million documents	Active
CiteSeerX	A	1998–current	Over 2 million documents	Active
Science Direct	A	1997–current	Over 11 million articles	Active
DBLP	A	09.2006–current	Over 2.3 million articles	Active
ACM DL	A	1997–current	391,000 full text articles	Active
Mendeley	A	07.2013–current	Over million documents	Active
Delicious	NA	09.2003–current	Around 1 billion links	Active
YouTube	NA	02.2005–current	Over 1 billion visitor users/Over 6 billion hours of video watching per month	Active
Flickr	NA	02.2004–current	87 million users/over 6 billion images	Active
43Things	NA	01.2005–current	Over 3 million users	Active
Pinterest	NA	03.2010–current	70 million users worldwide	Active
Facebook	NA	02.2004–current	1.2 billion	Active

can be retrieved at any time as needed: the data is updated in a system adaptive way, and the data in these services are mirrored in our proposed system with a frequency that can be determined adaptively. Table 1 explains the status of popular web-based academic and non-academic tools as of now.

The some of the services that are marked as *academic* and *nonacademic* in Table 1 are directly integrated into the our proposed system (citeulike, Connotea, GoogleScholar, GoogleScholar Advance, Microsoft Academic Search, Delicious) and the others are a natural candidate for implementing our algorithm. Note that Microsoft Academic Search has *active* status but its database is not updated since 2012.

As the web-based academic or nonacademic tools enabling storing, tagging and sharing documents have gained popularity, an emerging need has appeared for supporting these tools by using their existing services via Web Service wrappers with added capabilities. To address this challenges, an ideal architecture should meet the following requirements: a) *uniformity*: the architecture should support one-to-many services among information resources and their communication protocols; b) *federation*: the architecture should present a federation capability where different services belonging to different annotation resources on the web can interoperate with each other; c) *interoperability*: the architecture should be interoperable with different kinds of clients on the web; d) *performance*: the architecture should

search/retrieve/store metadata for scholarly publications with negligible processing overheads; e) *persistency*: the architecture should be able to back-up metadata about digital records without affecting the system performance; and f) *fault tolerance*: the architecture should be distributing metadata describing a digital content and managing redundancy of metadata about digital entities in acceptable rates. Fig. 1 illustrates a model of building a system hierarchy where search tools and existing services of social bookmarking tools can be used with added capabilities to collect and manage metadata and data for scientific content. Our goal is to define the practical extent of existing annotation tools for scholarly publications based on information retrieval and management in a consistent way.

We propose a Digital Information Service framework that improves the previous work [23] in great detail for reconciling distributed digital entities that addresses the challenges of discovering, retrieving, sharing and managing distributed data located on web-based systems in a Service Oriented Architecture where communications are provided through the Web Service technology. The proposed system provides users with ability to access their data even if a web-based system retires and discontinues its service in the future.

In this study, we present the semantics and the architectural design of the centralized Digital Information Service. We introduce a prototype implementation called IDIOM (Internet Documentation and Integration of Metadata) of

Download English Version:

<https://daneshyari.com/en/article/396663>

Download Persian Version:

<https://daneshyari.com/article/396663>

[Daneshyari.com](https://daneshyari.com)