Future Generation Computer Systems 27 (2011) 960–965

Contents lists available at ScienceDirect

Future Generation Computer Systems

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

Using ontologies for resource description in the CineGrid Exchange

Ralph Koning, Paola Grosso *, Cees de Laat

SNE group, Informatics Institute, University of Amsterdam, Science Park 107, 1098XG Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 17 March 2010
Received in revised form
23 July 2010
Accepted 29 November 2010
Available online 4 December 2010

Keywords:
Semantic Web
CineGrid Exchange
Network topology description
Services description

ABSTRACT

Moving media data between CineGrid Exchange nodes or external sites, and properly handling them at source and destination, is a challenge for the CineGrid community. The heterogeneous nature of the Exchange nodes, and in particular the lack of a globally accepted description mechanism for distributed media repositories, requires customized solutions to realize the real-time delivery of digital assets.

To address these issues we developed a Web Ontology Language (OWL) ontology, the CineGrid Description Language (CDL). CDL, coupled with the already existing Network Description Language (NDL), provides a common description of the various network and computing components. Tools for resource discovery, selection and composition will use CDL for data delivery, storage and viewing use cases.

We present here the CDL ontology and an initial demonstration of its applicability.

1. Introduction

The CineGrid Exchange "*is a distributed digital media repository designed to support CineGrid Member-driven testbeds for digital media asset management, distribution and preservation applications*" [1].

At present seven sites (*nodes*) comprise the CineGrid Exchange infrastructure:

1. Academy of Motion Picture Arts and Science (AMPAS) in Los Angeles;
2. California Institute for Telecommunications and Information Technology at the University of California San Diego (UCSD/Calit2);
3. Czech Republic National Research and Education Network in Prague (CESNET);
4. Electronic Visualization Laboratory at the University of Illinois at Chicago (UIC/EVL);
5. Keio University Research Institute for Digital Media and Content in Tokyo (Keio/DMC);
6. University of Amsterdam (UvA);
7. Ryerson University Digital Cinema and Advanced Visualization Laboratory, in Toronto.

The distributed nature of the Exchange repository is supported by software tools such as iRODS, used to store the content in the various locations [2]. iRODs is essential for the accomplishment of the asset management and preservation goals of the Exchange.

Our research is focused on the information framework needed for the distribution applications. Our challenge is to handle the heterogeneous hardware present at each site, the variety of distribution use cases (e.g. processing, storage etc.) and the different content formats, while hiding this complexity from the user.

The main research questions that motivated our work are:

- Which information is needed to identify, select and reserve the computing and networking resources in the Exchange?
- How do you build and instantiate a (dynamic) testbed suitable for the real-time delivery of CineGrid data?

We concluded that a unified resource description is crucial for discovering, selecting and reserving resources in CineGrid Exchange. In this paper, we present an ontology based resource description mechanism, the CineGrid Description Language (CDL). CDL is our answer to the first research question. The second research question is the subject of our future work.

The rest of the paper is organized as follows. First, we review the related work on using Semantic Web technologies in describing network resources; then we present the schema of the current version of CDL. Finally, we show a use case that demonstrates the usability of the CDL.

2. Related work

Tim Berners-Lee introduced the "Semantic Web" at the end of 1990s; the term became well known after he and his co-authors published an article in Scientific American in 2001 [3]. The idea proposed was that computers would be able to process information on the Web, and take higher level decisions, if the

* Corresponding author.
E-mail addresses: r.koning@uva.nl (R. Koning), p.grosso@uva.nl, grosso@science.uva.nl (P. Grosso), delaat@uva.nl (C. de Laat).

0167-739X/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.future.2010.11.027

information was presented to them in a suitable format. Since that time the Semantic Web has been widely adopted and it has grown enormously. It has proven to be suited for the categorisation of concepts within many different knowledge domains.

The information that computers and programs can consume is presented in so-called *ontologies*. Tom Gruber [4] gives a definition of ontology that is well accepted in the Artificial Intelligence and Semantic Web community. He defines an ontology as a formalisation of shared concepts; he calls it a description of the concepts and relationships that can exist for an agent or a community of agents. He makes the very important point that the main purpose of ontologies is that they enable sharing and reuse of knowledge.

Ontologies are expressed using the Resource Description Framework (RDF) [5,6], or in more recent years, by using the Web Ontology Language (OWL) [7,8]. Both RDF and OWL define *subject–predicate–object* triplets, used to describe a resource (the subject), its properties (the predicate) and the value of this property (the object). It is worth to observe that the object can be itself a resource, used as subject in other statements. This implies that ontologies build complex graphs of connected knowledge.

We had already applied Semantic Web when we developed the topology descriptions of optical networks. Optical networks are circuit-switched, and network paths (*lightpaths*) are built dynamically at the time applications request them [9,10]. Lightpaths cross administrative domains and to create them several pieces of information are necessary: for example, who is authorized to use resources in a domain, which devices are present and which configurations are possible. We defined for this purpose the Network Description Language (NDL) [11,12].

NDL uses RDF to define a series of schema, which provide the building blocks to describe optical networks. Several large education and research networks have adopted NDL and included it in their network management applications. We showed that NDL allows the exchange of topology information in real networks; we also demonstrated that path finding applications can consume NDL information to determine the existence of lightpaths between end points [13,14]; NDL can be further manipulated to create abstracted view of networks, useful for inter-domain topology exchanges [15].

Our experience with NDL showed us how useful a common information system is, when different domains need to interoperate. This is also, as we said before, the challenge in the CineGrid Exchange. Therefore, the step to the development of an ontology, an information model, that would describe CineGrid resources was obvious.

3. CineGrid Exchange applications

Testbeds involving nodes of the CineGrid Exchange have four main purposes:

- storing of data;
- streaming of data, e.g. sending data from storage to a remote node;
- displaying of data, e.g. sending data from a receiving node to video screens and projectors;
- manipulation of data, e.g. transcoding, color correction and resizing applications.

A common functionality that the testbeds must support is the transport of data, through network elements.

In order to make dynamic decisions, we need to know, in real time, the physical and logical relations and links between elements within the Exchange and their supporting networks. For example, given a certain destination display, we will need to decide where to stream data from and if it is necessary to transcode the data along the way.

To accomplish this, we need topology information to:

- produce up-to-date maps of the connections between elements at different nodes, the inter-domain links;
- produce up-to-date maps of the connections between elements at one node, the intra-domain links.

We also need to know which services are provided by all elements and their properties, such as:

- type of supported services;
- formats it can accept or produce;
- required bandwidth.

With this information the distribution applications can compute a number of paths that make efficient use of compatible resources; combined with real-time usage statistics of the elements the application can narrow this wider list down to the best available paths.

We created CDL to capture all this information, and to describe it in a machine readable format.

4. CineGrid Description Language

The CineGrid Description Language [16] provides a way to describe the resources and services used within the CineGrid Exchange. As we had done with NDL, we also decided to build CDL on Semantic Web principles.

A first important aspect is that the information can be distributed. Semantic Web allows different sites to publish and maintain their own descriptions and to point to peers where appropriate. Administrators of CineGrid Exchange nodes can maintain information independently, and still be fully interoperable.

Semantic Web also allows to point to more (and different) information from other sources e.g. websites, documents, RDF descriptions and other ontologies. Material in the CineGrid Exchange will have different origins, and we can expect that the original producer will maintain some information on it outside the Exchange. Another obvious reusability motivation for us is the description of the network resources in NDL.

Finally, Semantic Web provides a formal and strict way to represent information, which makes it easy to be processed by computer systems. We choose OWL as our syntax, where NDL uses RDF. Our main motive was that OWL provides more structures to perform logical reasoning.

The philosophy behind CDL is to define and categorise the concepts that are used in CineGrid use cases. CDL is divided in two parts, an infrastructure ontology and a service ontology. The service ontology describes the tasks a device can perform for the users of the Exchange. Devices in the Exchange nodes perform multiple types of tasks, possibly at the same time. We map these tasks into services; and the user of the ontology deals directly with services.

In order to do resource planning the service ontology had to be layered onto NDL. This seemed sufficient at first, but ultimately we were not able to map the concepts of node and exchange to NDL without modification. Because we needed these concepts, we decided to create our own infrastructure ontology. This also provides us with a clean separation between the languages, and removes the dependency on each other. Additionally this gives us the flexibility to use other schema for network descriptions, like for example the Network Markup Language currently under development in the OGF [17].

Fig. 1 shows the classes used in CDL and their relations to each other. The classes related to the infrastructure are marked yellow and the classes representing the different services are marked green.

Let us first look at the services part of the ontology. We have a generic `Service` class. We define three main services that