Chapter XII
A Tutorial on RDF with Jena

Wan-Yeung Wong
The Chinese University of Hong Kong, Hong Kong, China

Tak-Pang Lau
The Chinese University of Hong Kong, Hong Kong, China

Irwin King
The Chinese University of Hong Kong, Hong Kong, China

Michael R. Lyu
The Chinese University of Hong Kong, Hong Kong, China

ABSTRACT

This chapter gives a tutorial on resource description framework (RDF), its XML representation, and Jena, a set of Java-based API designed and implemented to further simplify the manipulation of RDF documents. RDF is a W3C standard which provides a common framework for describing resources in the World Wide Web and other applications. Under this standard framework with the Jena, different resources can be manipulated and exchanged easily, which leads to cost reduction and better efficiency in business applications. In this tutorial, we present some basic concepts and applications of RDF and Jena. In particular, we use a television object to illustrate the usage of RDF in describing various resources being used, the XML syntax in representing the RDF, and the ways Jena manipulate various RDF documents. Furthermore, complete programming codes with detailed explanations are also presented to give readers a better understanding of Jena. References are given at the end for readers’ further investigation.
INTRODUCTION

The resource description framework (RDF) (W3C, 2004a) is a W3C (W3C, 2005b) standard which is commonly used to describe resources for any application. For business applications, a resource may be a product, a service, or a person. Since having a standard framework to manipulate different resources often leads to cost reduction and better efficiency, RDF is widely used in business applications. In this chapter, we give a simple tutorial on RDF, the language (RDF/XML) used by RDF, and a Java API (Jena) for manipulating RDF/XML. We assume that readers should have some background knowledge on URI, XML, and Java. Readers may refer to Wu (2004) for more information about programming in Java.

We mention that RDF is used to describe resources such as products, services, or people. It provides the data model and XML (W3C, 2005a) syntax so that RDF documents can be easily exchanged by different applications. The XML language used by RDF is called RDF/XML. Moreover, the use of RDF/XML makes RDF documents to be both human readable and computer readable. There are some examples for using RDF such as RDF Site Summary (RSS) (RSS-DEV Working Group, 2000) and Friend of a Friend (FOAF) (Brickley, 2005). The former one is designed for the Web syndication, while the latter one is designed to describe people, interests, and interconnections.

The first work on RDF was started by R. V. Guha when he was with Apple Computer and later with Netscape. In 1999, the specification of RDF data model and XML syntax was published in W3C. The work continued and a new specification of RDF was published in 2004, completely replacing the old specification rather than being assigned a new version number.

In this tutorial, we first use a television product as a resource example. In RDF, we use a URI to uniquely identify a resource (The Internet Society, 2005). The URI of a television resource may consist of the company information (e.g., www.kingstv.com), category (e.g., plasma), and model number (e.g. PSM2000) like “http://www.kingstv.com/plasma/PSM2000”. Unlike the URL of a Web site, the URI of a resource is not necessary to be Internet accessible. For example, the aforementioned URI that begins with “http” does not necessarily have to represent that the television resource is accessible via HTTP. Such URI only denotes the abstract notion of world peace.

After defining the television resource, we can describe the resource by its properties and property values. A property is a resource that has a name, while a property value is the value of the property. For a television resource, the properties may be its model number, detail description, and price. Similar to the resource identification, we use a URI to uniquely identify a property. For example, the URI of the model number property may be “http://www.kingstv.com/tv_property/model” and its property value may be “PSM2000”. In addition, a property value can be another resource. We may have a user property which value is a resource for describing a person who uses that television.

The whole scenario can be represented by a labeled, directed graph called RDF graph (see Figure 1). Inside an RDF graph, we use an ellipse to represent a resource, an arrow to represent a property, and a rectangle to represent a non-resource property value (literal). For a literal (rectangle), further properties are not allowed. On the other hand, further properties are allowed for a resource (ellipse). In this example, the user resource has no further property although further properties are allowed. Besides the RDF graph, the whole scenario can be represented by a list of RDF statements. An RDF statement is a triple that contains a subject (resource), a predicate (property), and an object (property value). Table 1 shows a list of RDF triples, which is equivalent to the RDF graph in Figure 1.

Figure 2 shows a more complicated example of an RDF graph. The television resource contains two properties “model” and “accessories”. The