

RDF 原理、结构初探

姜恩波

(中国科学院成都文献情报中心 成都 610041)

【摘要】 RDF 是一个综合多种元数据标准来描述资源的框架体系, 极具发展潜力。它可以应用到搜索引擎、电子商务、数字图书馆信息等广泛的领域中。本文介绍了 RDF 的功能、目标和三个组成部分各自的作用。

【关键词】 RDF 资源描述框架 元数据

【分类号】 TP391

RDF Structure and Principle

Jiang Enbo

(The Chengdu Documentation and Information Center of CAS, Chengdu 610041, China)

【Abstract】 RDF is a framework which can use multiple metadata to describe the net resource. RDF will be applied into wide areas such as: search engine content description and DL et. al.. This article is to introduce the target and function of RDF and the three element of RDF: RDF data model, RDF schema and RDF syntax.

【Keywords】 RDF Metadata

1 RDF 概述

Resource Description Framework (RDF) 是 W3C 于 1999 年颁布的一个因特网建议。它的功能是利用当前存在着的多种元数据标准来描述各种网络资源, 形成人机可读, 并可以由计算机自动处理的文件。RDF 的目标是建立一个供多种元数据标准共存的框架。在这个框架中, 能够充分利用各种元数据的优势, “并能够进行基于 Web 的数据交换和再利用”。因此, RDF 的关键是框架结构。

元数据, 是“关于数据的数据”(Data about data)。它是相对于对象数据而言的, 是关于对象数据的一种概括性、实质性的描述。MARC 数据就是一种元数据。近年来, 随着因特网的发展和信息的丰富, 出现了多种元数据标准(格式)。如 DC (Dublin Core)、PICS (Platform of Internet Content Selection)、WF (Warwick Framework)。这些元数据对网络资源进行描述、组织和整理, 使之有序, 方便利用。用户不必直接接触对象数据源就可以决定取舍。规范化的元数据描述对网络信息的组织、挖掘、检索和利用十分有益。

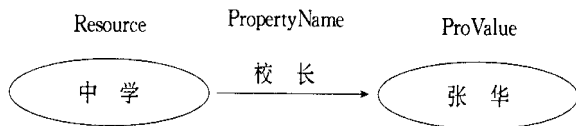
鉴于各种元数据各自发展, 优势无法综合利用, 并且, 内容互有重复的状况, W3C 成立了 W3C Resource Description Framework 工作组, 负责研究并提出一个综合性的框架来解决这个问题。因此, RDF 可以看作是一个元数据的“容器”。

RDF 框架由三个部分组成: RDF Data Model、RDF Schema 和 RDF Syntax。Data Model 形成对资源的形式描述; Schema 定义描述资源时需要的属性类及其意义、特性; Syn-

tax 则把形式描述通过其宿主语言 XML 转换成机器可以理解和处理的文件。

2 RDF 三个组成部分 (RDF Data Model)

RDF 提供的一个简单但功能强大的模型 (Model) 通过资源属性及其相应值来描述特定资源。模型定义为: 它包含一系列的节点 (N); 它包含一系列属性类 (P); 模型是一个三元组: {属性类 节点 节点或者原始值 (字符串或数值等) V}; 每个 Data Model 可以看成是一个实体——关系图。模型中所有被描述的资源以及用来描述资源的属性值都被看成是“节点” (Node)。由资源节点、属性类和属性值组成的一个三元组叫做 RDF 陈述 (Statement)。陈述可以理解为“资源 R 具有值为 V 的属性 P”。在一个模型中, 陈述即可以作为资源节点, 同时也可以作为值节点出现。所以一个模型中的节点有时不止一个。这时, 用来描述资源节点的值节点本身还具有属性类和值, 还可以继续细化。举个例子, 陈述“张华是这所中学的校长”。中学是被描述的资源, 它的一个属性是“校长”, 而这个属性的值是“张华”。这里“张华”, 节点可能还有一些属性, 如年龄、性别等, 可以继续对“张华”节点进行描述。下面是用实体——关系图来表示的这个陈述。



2.1 RDF Schema

“Schema”的功能象一个字典, 可以理解为大纲或者规范。RDF schema 的作用是:

收稿日期: 2001- 04- 28

- (1) 定义资源的属性类、语法、属性值的类型;
- (2) 定义资源类 (Type or class) 以及属性所应用到的资源类;
- (3) 申明 (Declare) 由一些机构定义的元数据标准的属性类。

RDF Schema 使用一种机器可以理解的体系来定义描述资源的词汇。

RDF Schema 的定义和宣布使用了 XML Namespace (命名空间) 机制。XML Namespace 的作用是用来避免不同元素被命名为相同的控制标记而给应用带来的困扰。在实际应用中有些相同名称的控制标记在不同的领域和机构中的含义是不同的。而 XML 是非常注重控制标记的语义含义的, 所以当不同应用领域中使用了同名称的控制标记时, XML 采用了这一方法来区别之。在 RDF Schema 中, 每个 RDF 资源都有一些核心类和属性。核心类包括:

RDFS: Resource: 任何利用 RDF 来表示、描述的事物都被称为“资源” (Resource)。

RDFS: Class: 这里 Class 类似面向对象中“类”的概念, 指的是事物的一类。这一类中的一个具体事物叫做“实例” (Instance)。比如, 哺乳动物和大象的关系。

RDFS: Property: 指资源的属性。

核心属性有 rdfs: type、rdfs: subclassof、rdfs: seeAlso 等。

2.2 RDF Syntax

RDF Data Model 提供了一个抽象的框架来描述资源。Schema 定义和声明了一些属性类。但是, 要形成人机可读的具体文件并利用计算机处理, 则还需要一个具体的语法体系来完成。RDF 以 XML 为其宿主语言, 通过 XML 语法实现对各种元数据的集成。下面的例子是通过 DC 和 ISO8601 来描述一本书的信息。

```
< ? xml version = "1. 0 " encoding = "gb2312 " standalone =
"no " ? >
< ? xml : : namespace href = "http: // www. w3. org/ RDF/ RDF/
" as = "RDF " ? >
< ? x ml : : namespace href = "http: // purl. ocl c. org/ DC/ " as
= "DC " ? >
< ? x m l ns = "http: // iso. org/ ISO8601/ " >
< RDF: RDF >
< RDF: Description RDF: HREF = "http: // uri- of- Document
- 1 " >
< DC: Creator> 王伟亮, 张力< /DC: Creator>
< DC: Title> VB 网络编程入门< /DC: Title>
< DC: Publisher> 北方大学出版社< /DC: Publisher>
< DC: Format> text/text< /DC: Format>
< DC: Date>
```

< ISO8601> 2000- 08- 30< /ISO8601>

< /DC: Date>

< DC: Rights> Copyright 北方大学出版社版权所有< /DC: Rights>

< /RDF: RDF>

< ? xml version = " ? " encoding = " ? " standalone = " ? " > 是每个 XML 程序所必需。它声明了版本, 编码体系 and 是否与其它文件相链接。接下来定义了三个命名空间。命名空间定义的格式是

< ? namespace href = "some- uri", as = "abbreviation" ? >

命名空间的定义有内定声明 (Default Declaration) 和明确声明 (Exp; ocot Declaration) 两种方式。前两个使用的是明确声明, 所以, 在不同元数据标准的属性类前要加上相应的标识, 如 "DC: Creator", 理解为 "DC 命名空间中的属性元素 Creator"。而 ISO8601 声明则使用的是内定声明。

对于一个 RDF Data Model, RDF 规范有两种方式来编码。一种是简略语法; 另一种则是完整语法。它通过正则规范来表达详细的 RDF Data Model。显见, 上述例子使用了完整语法。因为 RDF 是 XML 的一个应用, 所以, RDF 文件也可以不需要文档定义说明 (DTD)。但是它必须是 Well-Formed, 这也是一个 XML 文件最基本的规格限制, 否则无法通过语法解析器。

3 RDF 规范的应用前景

虽然 RDF 规范还没有成为因特网标准, 描述它的文档也随时可能更改。但是, 毫无疑问, 它有着广泛的应用前景。

(1) RDF 的基本功能是描述网站、文献等网络资源, 这将使得它在因特网范围内得以广泛地应用;

(2) RDF 为多种元数据描述资源提供了一个框架, 综合了多种元数据标准的优势;

(3) 利用元数据描述对象数据, 可以减小实际应用中网络的传输量, 节省网络带宽, RDF 的功能和目标符合这一要求;

(4) 用户通过规范的元数据内容就可以决定对象数据的取舍, 节省时间和精力;

(5) RDF 以 XML 为其宿主语言。XML 摒弃了 SGML 的缺点, 显得简略、清晰且应用方便, 是目前最具发展前景的标记语言。

参考文献:

- [1] Dan Brickley, R. V. Guha : Resource Description Framework (RDF) Schema Specification 1.0
- [2] Eric Miller: An Introduction to the Resource Description Framework
- [3] 程爱变. 试论资源描述框架 (RDF) —— 一种极具生命力的元数据携带工具. 现代图书情报技术, 2000, (6): 62

启 事: 为适应我国信息化建设需要, 扩大作者学术交流渠道, 自 2001 年起本刊加入《中国学术期刊 (光盘版)》和“中国期刊网”, 其作者文章著作权使用费交中国版权保护中心统一处理, 如作者不同意将文章编入该数据库, 请在来稿时声明, 本刊将做适当处理。