

文献综述

题 目 RDF 数据检索处理技术综述

学生姓名 谢先斌

专业班级 计算机科学与技术 10-01 班

学 号 541007010144

院（系） 计算机与通信工程学院

指导教师(职称) 王岩（讲师）

完成时间 2014 年 3 月 14 日

RDF 数据检索处理技术综述

1. 前言

随着语义网的快速发展，目前 Web 上语义网数据已经达到相当的规模，成为重要的信息和知识来源。因此，RDF 数据浏览^[1]的研究开始得到广泛关注。通过对比传统数据检索和 RDF 数据检索两种常见的检索方式，得出 RDF 数据检索具有快速、准确的特点。基于人们具有获取准确信息的现实需求，因此，研究基于 RDF 的数据检索具有重要的现实意义。

2. 语义网的起源和发展

2000 年 12 月，万维网创始人 Tim Berners-Lee 在 XML2000 会议上正式提出了下一代互联网的概念--语义网^[2] (Semantic Web)，目的在于赋予万维网上所有资源一个唯一标识，并且资源之间可以建立机器可处理的各类语义联系^[3]，从而使资源之间的关系更加明确，对资源的搜索更加精确。

语义网是对当代万维网的延伸和扩展，它使用一种明确的、形式化的方式来表示资源信息，有效的提高了异构系统之间的有效检索和访问。

语义网是人工智能领域和 Web 技术相结合的产物，人工智能领域中的知识工程研究从孤立的知识库系统逐渐发展到基于 Internet, Ext ranet 的信息系统集成，最后扩展到整个 Internet。在这个研究过程中，逐渐加深了对知识表达和推理的认识，并总结出了一些新的描述和推理方法。另一方面，万维网经过十几年的发展，积累的海量数据需要一种新的、机器可以自动完成的方式来处理和管理^[4]。因此，当两个领域的积累都比较成熟，而且有了需求，就必然水到渠成。

目前，国内外关于 RDF 的研究十分火热，基于 RDF 数据检索的研究主要分为以下几种：

DBpedia^[5]是从维基百科中提取结构化信息，并使这些信息在互联网上可共享使用的网站。

基于 RDF 的科技论文搜索引擎的设计与实现，提出一种基于 RDF 的科技论文搜索引擎的设计与实现方案。

HealthCyberMap 是一个医学语义 Web 项目，基于修饰词的都柏林数据集描述 RDF 元数据。它使用 Protégé-2000 来进行数据建模，并使用相关技术连接美国国家医学图书馆的知识资源数据库。

3. RDF 数据检索和传统检索的对比

当前因特网在信息表达和检索方面存在严重的缺陷，例如，对“UEFI”这个概念进行检索，传统的搜索引擎仅能搜索到含有 UEFI 关键字的信息，但并不一定是具体的概念或介绍。虽然这些信息用户可以直接阅读处理，但计算机不能读取具体的语义信息，因此限制了计算机智能处理信息的能力，不适合对信息进行准确的定位。而实现信息资源的语义检索，是提高数字化信息资源的重要突破方向，也是信息搜索领域面临的重要挑战。

目前互联网信息的传递以 HTML 网页为主体，大量的信息信息的表现形式并没有和内在结构（HTML/CSS 样式表）等分离，非常不利于计算机的自动化处理。Tim Berners-lee 在 2000 年提出了下一代万维网的构想--语义网（Semantic Web）。语义网网页所包含的语义信息支持计算机完成信息的推理和实现进一步的自动化处理。Tim Berners-Lee 对语义网的描述是：语义网并非一个从无到有、孤立发展的万维网，而是对万维网的扩展和延伸^[6]，语义网信息具有定义良好的含义，计算机可以凭借概念的定义声明和逻辑推理规则寻找到资源对象的含义，从而使得人-机之间和机-机之间合作处理更加有效；数据通过语义网中的定义和链接可以更有效的实现各种检索、重用和集成。

4. RDF 数据检索可行性分析

下面对基于 RDF 的数据检索做技术可行性、经济可行性和操作可行性三个方面进行分析。

4.1 技术可行性

RDF 和 SparQL^[7]是语义网检索的基础，课程知识库系统可利用惠普实验室的开放

源代码 Jena Semantic Web Framework 在 Java 应用程序中搭建 RDF 数据模型，完成系统对 RDF 文件的操作需求，加以老师的指导和帮助可以解决该技术问题。

利用 SSH^[8]框架技术可以快速搭建基于 B/S 架构^[9]的先进、安全和跨平台的 JAVA WEB 系统，利用 CAS^[10]可以保障系统各项目之间的安全，应用也不是问题。

DES 算法^[11]是常见的加密算法，该应用主要难点是加密模型的建立，网上有很多该应用实例，使用起来也不是问题。

综上技术可行性分析，课程知识库系统不存在技术性问题。

4.2 经济可行性

由于课程知识库系统的主要应用背景是毕业课程设计，其实并不注重直接的经济效益，其实现目的是针对毕业生自身水平和能力的提高，并且程序发布经济要求不高，只要有一台可以运行 JDK、Tomcat 和 Nginx^[12]的服务器即可，对系统环境和配置并没有刚性的需求，所以不用考虑该系统的经济问题。

4.3 操作可行性

课程知识库系统主要面向的使用对象是学生等对课程知识点进行精确、快速的查询，只要用户熟悉常用的搜索引擎如百度、Google 等就可以使用该系统，并且由于系统采用 B/S 架构，客户端只需要一个浏览器即可访问系统。因此，用户只要会一些电脑操作，就可以进行查询所需的知识点，而不需要用户有较高的计算机专业知识。该系统对于网站的维护也十分方便，系统管理员只需登录系统后台就可以对系统信息进行维护。

5. 语义网研究的前景

万维网是图像与文本的数据库，语义学引入后将会从根本上改变万维网的性质，将信息进行解释、交换和处理，从搜索服务器端通过分散于各处理服务器中的应用分析来搜索数据库，从多种来源可以收集一些机器可读数据，处理并推理出新的事实。语义网可以使分散于全球成百万的独立数据库融合，最终使用户独立运用 Internet 上庞大的信息资源。目前主要应用于智能信息检索、基于 A-gent 分布式计算^[13]、Web Service^[14]和企业数据管理等方面。

如今，对 **Ontology** 语言^[15]、推理的研究成果越来越多，也有不少实际应用的实例出现。随着研究的不断深入，将会发现语义网不仅仅是执行单个任务的工具，如果设计合理，语义网还能从整体上帮助人类知识的发展。因为在语义网中，每个概念都由 **URI** 定义，任何人花最少的工夫就能表达他们发明的新概念。语义网的统一逻辑语言将这些概念连接到全球的 **Web** 上。语义网所具有的广泛性和分散性特点可以帮助用户将人类的知识和工作发布给软件代理，并进行有意义的分析和推断，从而为人类提供更全、更新、更快、更准的知识服务。

参考文献

- [1] 吴鸿汉,瞿裕忠. RDF 数据浏览的研究综述[J]. 计算机科学,2009,02:5-10+41.
- [2] 李洁,丁颖. 语义网关键技术概述[J]. 计算机工程与设计,2007,08:1831-1833+1836.
- [3] 高继峰. 基于语义 Web 的智能信息检索系统研究[D].郑州大学,2006.
- [4] 崔亚辉. 语义网的起源与发展[J]. 机械管理开发,2009,05:186-187.
- [5] 朝乐门,张勇,邢春晓. DBpedia 及其典型应用[J]. 现代图书情报技术,2011,03:80-87.
- [6] Paula Jacobs ,文广. Tim Berners-Lee 编织一个还算简单的复杂 Web 网[J]. 今日电子,1995,06:47-48.
- [7] 陈彦. 应用 SPARQL 进行 RDF 数据查询[J]. 企业技术开发,2007,07:6-10.
- [8] 郑晨. 基于 SSH 框架的图书馆管理系统分析与设计[D].云南大学,2012.
- [9] 蔡长安,王盈瑛. C/S 和 B/S 的模式的选择和选择[J]. 渭南师范学院学报,2006,02:47-50+74.
- [10] Yale University. Central Authentication Service project[EB/OL]. [2014-05-29]
<http://www.jasig.org/cas>
- [11] 李联. 信息安全中的 DES 加密算法[J]. 现代电子技术,2005,09:118-120.
- [12] 杨伟强. 安全高效的 Nginx 反向加速技术在站群中的应用[J]. 信息系统工程,2013,03:83-84.
- [13] 骆正虎,杨敬安,骆祥峰,郑淑丽,张浩. 基于移动 Agent 的分布式计算模型研究[J]. 小型微型计算机系统,2002,03:300-304.
- [14] 王楠,刘心雄,陈和平. Web Service 技术研究[J]. 计算机与数字工程,2006,07:88-90+100.
- [15] 杨力. 从 RDF,DAML+OIL 到 OWL——Ontology 语言比较[J]. 农业图书情报学刊,2005,11:110-112+119.