本体与WEB语义检索

赵猛

(大连海事大学 信息与科学技术学院, 大连市 116026)

摘 要 目前，我们进入了信息爆炸的时代，万维网是为我们提供了丰富内容的信息源。但是传统的信息检索技术不能对这些信息进行有效的分类，所以，本文在现有的语义检索的基础上，以本体为依据，对语义检索进行简单分析与研究。

关键词 信息爆炸；本体；语义检索

**Ontology and Web Semantic Retrieval**

Zhao Meng

（Dalian Maritime University Department of Information and Science , Dalian 116026）

**Abstract** At present, we have entered the era of information explosion, the web provides a rich source of information content for us. But the traditional information retrieval technologies can not classify the information, therefore, based on the existing semantic retrieval, ontology as the basis, this paper makes a simple analysis and research on semantic retrieval.

**Key words** Information explosion; Ontology; Semantic Retrieval

**1**引言

WEB因其丰富的信息内容和访问的便捷性，使得网络成为能自动处理信息并且动态显示数据的多功能平台。但是现在的搜索引擎多采用字符匹配的算法，在用户进行搜索时，会存在许多无用的信息。针对目前检索方面存在的弊端，Tim Berners-Lee提出了语义Web的思想[1]，对现有Web进行扩展，对信息嵌入机器可阅读的，代表某类知识的标注，使计算机和用户都能理解，从而使计算机和用户能协同工作，这种理论的提出，为解决目前所遇到的问题提供的新的思路。

**2**本体(Ontology)与语义检索

**2.1** 本体的定义

本体出自哲学的概念，是关于存在的学说，反映其本质和规律的内涵。在人工智能界，Neches等人将Ontology定义为“给出构成相关领域词汇的基本术语和关系，以及利用这些术语和关系构成的规定这些词汇外延的规则的定义”[2]。Studers认为本体是共享概念模型的明确的形式概念说明，包含四层含义：概念模型，明确，形式化，共享[3]。

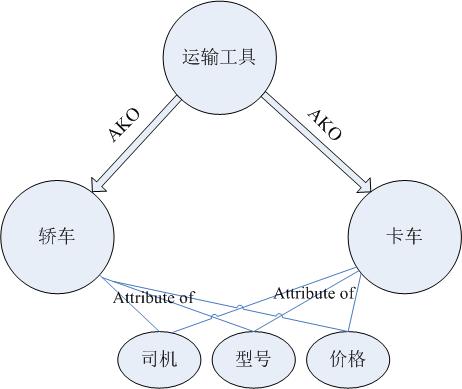


图1 简单本体举例

构造本体可以实现某种程度的知识共享和重用，以及提高系统通讯、互操作、可靠性的能力，本体允许使用一系列丰富的结构关系和非结构关系，如泛化、继承、聚合和实例化，并且可以为软件应用程序提供精确的[领域模型](http://baike.baidu.com/view/757895.htm)。例如，本体可以为传统软件提供[面向对象](http://baike.baidu.com/view/125370.htm)型系统的对象模式（object schema），以及类的定义。

**2.2** 什么是语义检索？

我们广泛使用的信息检索方法或者搜索引擎，都是以关键词匹配为基础的，这就使得检索的内容可能只符合关键字字面内容，却与用户的真正需求相差甚远。而语义检索则是基于概念匹配的检索方法，把关键字替换为含有语义的概念，从而使检索提升到概念级的检索，可以有效的去除无意义的关键字干扰，将检索结果通过语义来达到更加符合用户需求的目的[4]。

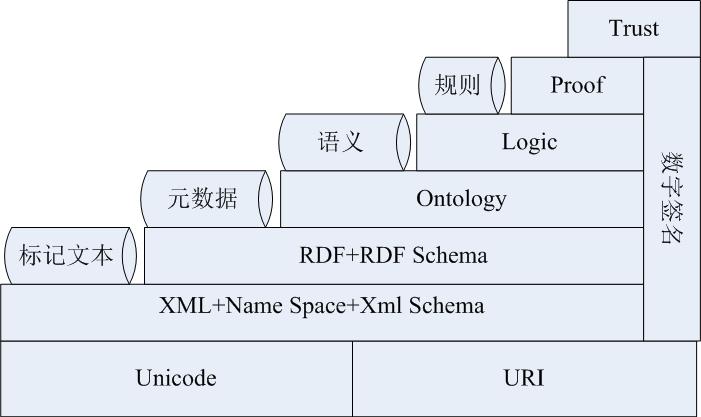


图2 语义网分层结构

**2.3** 资源描述框架(RDF)

资源描述框架是一个用于表达关于万维网上的资源的信息的语言。为了让Web信息可以被计算机自动处理，需要采用元数据来索引网上的信息，而资源描述框架就是专门用于表达关于Web资源的[元数据](http://baike.baidu.com/view/107838.htm" \t "http://baike.baidu.com/_blank)，RDF提供了一种用于表达这一信息、并使其能在应用程序间交换而不丧失语义的通用框架。

**3**基于本体的语义检索基本思想

最早在1994年，Voorhees就提出基于本体的查询扩展[4]。2004年，Navigli提出了基于本体注释的查询扩展方法，使用了WorldNet中的概念并对其进行扩充了注释[4]。经过不断的发展与探索，研究人员意识到要想更好的使用语义，就应该充分利用到本体中的属性和其他关系，将文档和用户查询中所包含的有价值的信息最大程度的提取出来，从而提出了基于本体的语义检索模型。

**3.1** 本体的构建

首先确定本体的领域，对本体所要完成的目的确定下来。接着应该列举出其中的概念，然后将这些概念按照一定的逻辑规则进行分组，并从其中通过固定的评测手段，选出较为关键的术语，以此来缩小本体架构的范围。最后建立本体的评价标准，将符合标准的本体以文件的形式存放[5]。

**3.2 语义检索**

语义检索就是将用户输入的语句，将语句中的词汇依据不同的属性分为不同的实体。然后从Web界面中提取出RDF层本体，并通过特定的处理机机制来提取出RDF在网页中嵌入的元数据。最后综合相似度与相关度，找到满足条件的结果，根据概念相关匹配度进行排序，最后反馈给用户。

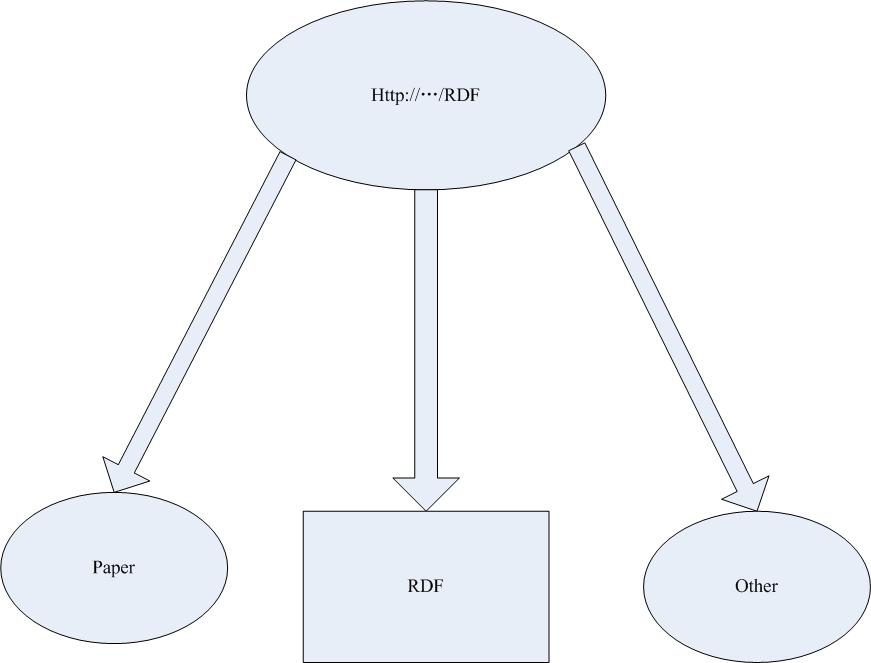


图3 提取的RDF本体层举例

其中提取的本体层就如上图所示，包含其类型和多个属性值的内容，并且可以存在指向其他资源的属性。

**3.2 结果评价**

语义检索的效果最终还是需要通过查准率(precision)和查全率(recall)来进行评价。





查准率和查全率的全面提高，就表明该检索方式能有效提高检索的效率。

**4**总结

总体来说，基于本体的语义检索相对于目前广泛使用的文档式检索，有效的提高了检索的效率，将用户输入的内容中的信息最大程度的利用起来，进一步的把语义检索的作用发挥出来。但是目前的语义检索在处理语义相关资源时，会将大部分资源用作寻找语义相关资源，导致检索会耗费大量的时间，减低用户友好性。同时，这种检索也会出现曲解语义的问题。因此，基于本体的Web语义检索仍然需要继续研究与探索，达到更好的搜索结果。

参 考 文 献

[1] Berners-Lee T, Hendler J. The Semantic Web. Scientific American, 2001,258(5):34-37.

[2]Deng Zhi-hong,Tang Shi-wei. Review of Ontology Research, Universitatis Pekinensis, 2002 38(5):730-733

(邓志鸿,唐世渭。Ontology 研究综述,北京大学学报,2002 38(5) 730-733)

[3]Studer R, Benjamins V R, Knowledge engineering principles and methods. Data and Knowledge Engineering,1998,25(122):161-197.

[4]王进,基于本体的语义检索研究[博士学位论文], 中国科学技术大学.2006

[5]Chen Yong, Lin Shi-ping. Technology of Sematic Search based on Ontology, Computer Engineering and Application. 2006:78-80

(陈泳，林世平. 基于本体的语义检索技术, 计算机工程与应用. 2006 78-80)

[6]Jiang hua. Research and Implement of Semantic Search Technology Based on Ontology. New Technology of library and Information service.2008 4:39-41

[7]Wang Xiao-dong, Zheng He. Study of semantic information retrival model based on Ontology. Computer Engineering and Design.2008 29(11):2939-2941