基于本体的Web信息检索

安东

（大连海事大学，大连，辽宁 116026）

摘 要： 对现如今的互联网中出现的基于关键字词的检索出现的查全率（Recall）和查准率（Precision）的问题。在查找中会得到大量的无关数据。提出了通过本体论基础上的语义分析和逻辑推理，使得计算机能够理解用户的查询意图，明确信息的需求，这样可以保证信息检索的查全率，而且也可以保证信息检索的查准率，查询的结果可以令客户满意，避免冗余。

关键词：本体论；本体；信息检索；语义网

Web information retrieval based on Ontology

An Dong

【abstract】 Based on the key words in the retrieval recall now appears in the Internet (Recall) and precision (Precision) problem. In the search will be a large number of independent data. The paper presents the ontology semantic analysis and logical reasoning on the basis of which the computer can understand the user's query intention, clear information needs so, can guarantee the recall of information retrieval, but also can guarantee the precision of information retrieval, the query results can make customer satisfaction, to avoid redundancy.

【Key words】Ontology; ontology; information retrieval; Semantic Web

0 引言

随着信息的高速发展，计算机在人们的社会中越来越普及。广大群众对数字信息的需求变得越来越强烈，随着信息不断的发展，信息技术开始面临巨大的挑战。目前，信息检索的查全率（Recall）和查准率（Precision）难以令人满意，往往会得到很多无用的信息。在当今因特网的迅速发展，使得如何处理大型数据，如何检索大型数据，并能够为用户提供高效、准确的数据检索方法，成为了现如今一项重要的研究内容。

目前的Web信息检索方法是基于关键字词的搜索。通过这样的搜索其查全率和查准率都无法得到保证。为了满足这些需求，我们提出了本体（Ontology）作为一种能在语义和知识层次上描述信息系统的概念模型建模工具【[[1]](#endnote-0)】。本体具有良好的概念层次结构和对逻辑推理的支持，因而本体（Ontology）得到了很多研究领域的重视。

1 本体的相关概念

1.1本体的含义

本体出自一个哲学术语，表示客观现实的抽象本质。 Neches 最早将本体引入到人工智能中。1993 年，Gruber 给出 了本体的一个最为流行的定义，即本体是概念模型的明确的 规范性说明；之后，Borst 对此定义进行修改，提出本体是共享概念模型的形式化规范说明。Studer提出的认为“本体是共享概念模型的明确的形式化规范说明”【[[2]](#endnote-1)】。这个定义的说明体现了本体的4层含义：

1. 概念模型（conceptualization）通过抽象出客观世界中一些现象的相关概念而得到的模型，其表示的含义独立于具体的环境状态。
2. 明确（explicit）所使用的概念及使用这些概念的约束都有明确的定义。
3. 形式化（formal）本体是计算机可读的，技能被计算机处理。

4、共享（share）本体中体现的是共同认可的知识，反映的是相关领域中公认的概念集，它所针对的是团体而不是个体。

总之，通俗的来讲，本体的作用就是来描述某个或多个领域内概念以及概念之间的关系。在万维网中，我们可能会用不同的术语来表达相同的含义，或者一个术语含有多个含义。因此，消除术语差异是很有必要的。目前的解决方法是，建立一个公共的本体。在设计该领域的时候，我们都使用这个公共本体来进行规范。

2 基于本体的信息获取

基于Ontology的信息检索，构造Ontology的过程有很多的方法。没有一个标准的Ontology构造方法，期中最有影响的是Gruber在1995年文献中提出的五条规则【[[3]](#endnote-2)】：

( 1) 明确性和客观性, 即 Ontology 应该用自然语言对所定 义的术语给出明确的、客观的语义定义。

( 2) 完全性, 即所给出的定义是完整的, 完全能表达所描 述术语的含义。

( 3) 一致性, 即由术语得出的推论与术语本身的含义是相 容的, 不会产生矛盾。

( 4) 最大单调可扩展性, 即向 Ontology 中添加通用或专用 的术语时, 不需要修改其已有的内容。

( 5) 最小承诺, 即对待建模对象给出尽可能少的约束。

当前对构造 Ontology 的方法和方法的性能评估还没有一 个统一的标准, 我们采用 Uschold & King 的“骨架法”的思想来 构造面向特定域的 Ontologies, 其过程如图 1 所示。

确定关注的领域和任务

本体分析

本体的表示

不符合

本体的评价

本体的建立

符合

图1 “骨架法”构造本体的流程图

基于Ontology的信息检索框架如图2所示，其工作流程如下：

通过“骨架法”构造出本体，例如对“苹果手机”的查询，可以分解成为关键字集{苹果，手机}，结合语义网络进行逻辑推理，形成的查询计划中会有针对性包含了apple公司，苹果公司的相关成员，苹果公司的相关产品。这些信息没有在关键字“苹果手机”中出现。如果基于关键字的差早，就可能会有很多关于“苹果”、“手机”等字眼的搜索结果被显示到用户的面前，这样就可能造成很大的冗余。影响了查准率。这些很多都可能是用户在查找中并不关心的内容，这些用户并不关心的内容返回给了用户，影响了查准率。

苹果手机

苹果手机

苹果平板

苹果公司产品

苹果公司

苹果手表

组成人员

图3 Ontology的系统框架

3结语

本文描述了关于本体论、本体、信息检索和语义网的相关概念。本体最早的源于是哲学术语，演变到，被计算机所引入，应用本体把本体（Ontology）作为一种能在语义和知识层次上描述信息系统的概念模型建模工具。基于本体的Web信息检索，能够更好的将用户对信息检索更加精确化的查找需求。通过建立明确性和客观性、完全性、一致性、最大单调可扩展性的本体，建立基于语义本体的Web检索。与传统的基于关键字的检索相比，查全率和查准率都得到了提高，查询结果更为合理可用。

4参考文献

1. 王诚，张璟.基于语义的Web信息检索 .西安理工大学。 1001-3695( 2005) 08-0111-02 [↑](#endnote-ref-0)
2. LAN-MAN Standards Committee. IEEE STD 802.11. Wireless LAN Medium Access Control and Physical Layer Specifications[Z]. [2008-05-31]. http://standards.ieee.org/getieee802/802.11.html. [↑](#endnote-ref-1)
3. Gruber T R. Towards Principles for the Design of Ontologies Used for Knowledge Sharing[ J] . Int. J. of Human Computer Studies, 1995, 43: 907-928. [↑](#endnote-ref-2)