

●王翠波 张玉峰 艾丹祥 金燕

基于语义网的信息组织与智能导航研究*

摘要 语义网技术可以引导人们进行语义层次上的信息分类、信息标引、信息索引、信息整合等方式的信息组织,实现一个有序的信息空间。基于语义网的智能导航是利用语义机制进行导航。图3。参考文献7。

关键词 智能导航 语义网 信息组织

分类号 G354

ABSTRACT Semantic web technology can guide people to do information classification, information indexing, information integration and other kinds of information organization at the semantic level, and then realize a orderly information space. The intelligent navigation based on semantic web is to use the mechanism of semantic web for navigation. 3 figs. 7 refs.

KEY WORDS Intelligent navigation. Semantic web. Information organization.

CLASS NUMBER G354

目前,网络信息空间中数据,主要以非结构化和半结构化形式存在,网络信息资源仅实现词法层面的管理,缺乏语义描述,网络信息的语义理解一直是智能导航研究的瓶颈问题。如何设计一系列计算机可理解的语义表达形式并支持有效推理,是实现网络环境中智能导航的关键。语义网的出现为智能导航研究带来一线曙光。

1 关于语义网

由于因特网所采用的 HTML 语言仅仅用于页面信息的显示和布局,主要面向用户阅读,信息内容和信息表现形式相混合的特点使得计算机难以阅读和理解信息内容的实质。这个缺陷在很大程度上影响了人们获得有用信息的效率。针对这一缺陷, Tim Berners-Lee 于 1998 年首次提出了语义网的概念,并于 2001 年 5 月在 *Scientific American* 上发表文章,通过描述 Agent 如何在语义网环境中进行智能型信息获取,展示了实现语义网后的美好前景:在相关技术的支持下,网络信息空间成为了一个巨大的知识网络,机器可以自动理解信息内容并进行知识推理,为人们提供高质量、自动化的多元信息服务^[1]。

作为现有 Web 网络的延续,语义网主要添加了两点:一方面,对半结构化数据和元数据采用 RDF 作为通用数据表现形式;另一方面,通过采用本体对数据层中使用的元数据进行定义和关联,揭示出数据源的概念模型。这样,与传统 Web 网络相比,语

义网具有了机器可理解的语义。在此基础上,人们可以通过构建各种应用系统实现信息自动搜集、信息自动分析、信息自动过滤等信息自动化处理。

建立语义网应用系统需要一系列相应的开发工具。总的来说,这些开发工具分为 3 类:一是本体建立工具;二是基于本体的语义标注工具;三是支持本体整合、本体存储、本体解析和推理应用的工具。

随着语义网相关技术的发展与成熟,各种语义网应用正受到广泛关注。其应用有如下特点:

(1)应用的多维性。语义网的应用包括很多方面,从基于语义网的个人信息管理,到企业联合知识管理,这些应用都力图通过利用语义网技术提高管理的效能。

(2)应用领域的广泛性。语义网的应用涉及多个领域。例如,在教育领域、医学信息、Web 服务和电子政务中的应用等。

(3)语义网应用所提供的服务层次有所不同。有的只包含基于语义网的概念检索;有的可以提供基于多 Agent 的主动服务。

2 基于语义网的信息组织

对杂乱无序的信息空间进行合理有序地组织是解决信息迷航的关键。语义网技术可以引导人们进行语义层次上的信息分类、信息标引、信息索引、信息整合等方式的信息组织,实现一个有序的信息空间。

2.1 语义元素分析

语义元素是基于语义的信息组织的基础。语义

* 本文系教育部人文社会科学研究博士点基金项目“基于信息构建的智能导航机理研究”(项目编号 03JB870002)的成果之一。

元素分析主要从3个方面进行:语义概念分析、语义关联分析和语义表示分析。

概念是思维的基本形式。它反映了客观事物的一般的本质的特征。语言是思维的媒介,概念需要通过一定的字、词或短语等描述元素来加以表达。通常多个描述元素可以表达同一个概念,如“电脑”、“计算机”和“微机”,在此概念下,这些描述元素构成同义关系;同一描述元素在不同的语境中又可以表示不同的概念。概念可以是具体的世界万物,也可以是抽象的范畴。不同的范畴抽象程度形成不同层次的概念。如由词汇概念“树”表示的范畴中,“树”属于最上层的范畴,它包括“樱花树”、“梅树”等从属层范畴,而在“梅树”下又包括“粉红梅”、“白梅”、“红梅”等基本层范畴。

语义关联主要体现在语义内容关联和语义结构关联两个方面。认知语言学认为词语意义的解读是认知者心理运算的结果,跟认知者的概念结构密切相关,词语的意义依赖于认知者的概念结构,任何语义内容的分析都必须依靠对概念的分析来实现^[2]。任何概念都不是孤立存在的,概念与概念之间存在一定的关系,除了范畴抽象程度不同而形成的种属关系,还存在等级关系、同义关系、反义关系、整部关系、包含关系等,概念结点、概念结点之间的关系两者共同组成了概念空间。可以通过概念间的多种关系及其紧密度来反映信息对象空间的语义内容关联。在超文本信息网络中,节点和链作为定义超文本信息结构的基本要素,组建了类似于人类联想记忆的网状信息结构。作为节点存在的信息块之间有着语义内容上的关联,超文本信息结构因此表现出一定的语义结构关联。

语义表示主要是对信息对象的语义内容以及信息对象间的语义关系进行描述。由于语义内容主要以语义概念的形式来表示人们对事物共同特点的认识,同时通过概念关联来体现信息内容的语义关系,因此语义表示可以看成是对概念及其概念关联进行

描述。根据概念和概念关系定义的不同,存在多种不同的语义表示形式,如概念树、概念网络、本体等。

2.2 基于语义的信息组织方式

(1)语义分类。所谓语义分类,是将概念划分到各个类和子类,具体表现形式一般为树状结构的语义树。语义分类揭示了概念与其他概念之间存在的上下位层次关系。概念的语义分类是标明一个概念的语义属性的常用手段之一,概念模型中的每一个概念词应该有语义分类的属性^[3]。

不同的概念空间模型有不同的语义分类结构。当前语义分类的主要问题是语义分类难以全面,分类又有交叉,分类的标准也不统一等等。随着语义网研究的深入和普及,本体作为一种具有全球规范性、共享性特点的概念空间,可以通过映射、整合等手段为语义分类提供有效途径。

分类是人类认识和把握事物的基本思维方法。类,是具有某种共同属性的事物或信息的集合,是人类在认识事物过程中对类似事物归纳、聚合的结果。就信息来说,各类信息都具有多方面的内在联系,形成立体或网状的结构。当大量零散、无序的信息经过分类,纳入一个特定的分类体系后,就可以揭示它们在整个系统中的位置,展示它们之间的隶属关系、相关关系等,使信息空间成为一个有机系统^[4]。笔者认为基于语义分类的信息组织应该从两个方面加以考虑:首先是概念如何被划分到各个类和子类,即语义分类的确定;其次是如何根据已有的语义分类进行文档分类。

(2)语义索引。语义索引就是在概念空间的基础上构造具有网状结构的索引,其结点是从文档中抽取出来的概念,同一文档可由具有相关语义的多个概念进行索引。通过概念的共现分析建立语义索引,然后将语义索引保存到语义空间。借助语义索引网络,能快速、准确、全面地定位待检索的概念。语义索引功能框架如图1所示^[5]。

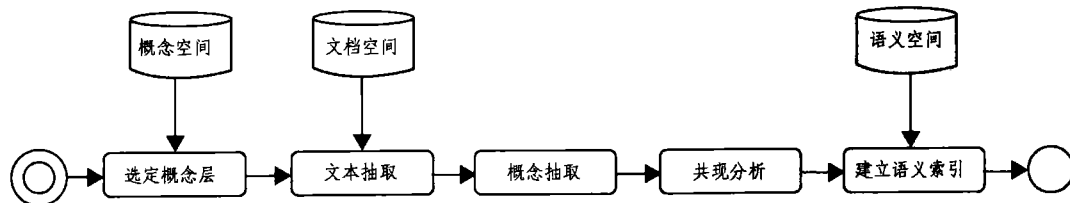


图1 语义索引功能框架

3 基于语义网的智能导航模型

基于语义网的智能导航是利用语义机制进行信息导航,以改变当前信息导航中缺乏语义的现状,提高导航效果。语义导航主要包括两个方面:一是对信息资源进行语义描述、组织和整合;二是考虑用户的

背景知识、信息需求等,对用户信息进行语义抽取,建立用户语义模型。在构建信息源语义模型和用户语义模型的基础上,导航系统采用一定的导航方法,把语义内容呈现给用户。图2显示了语义导航机制的模型。

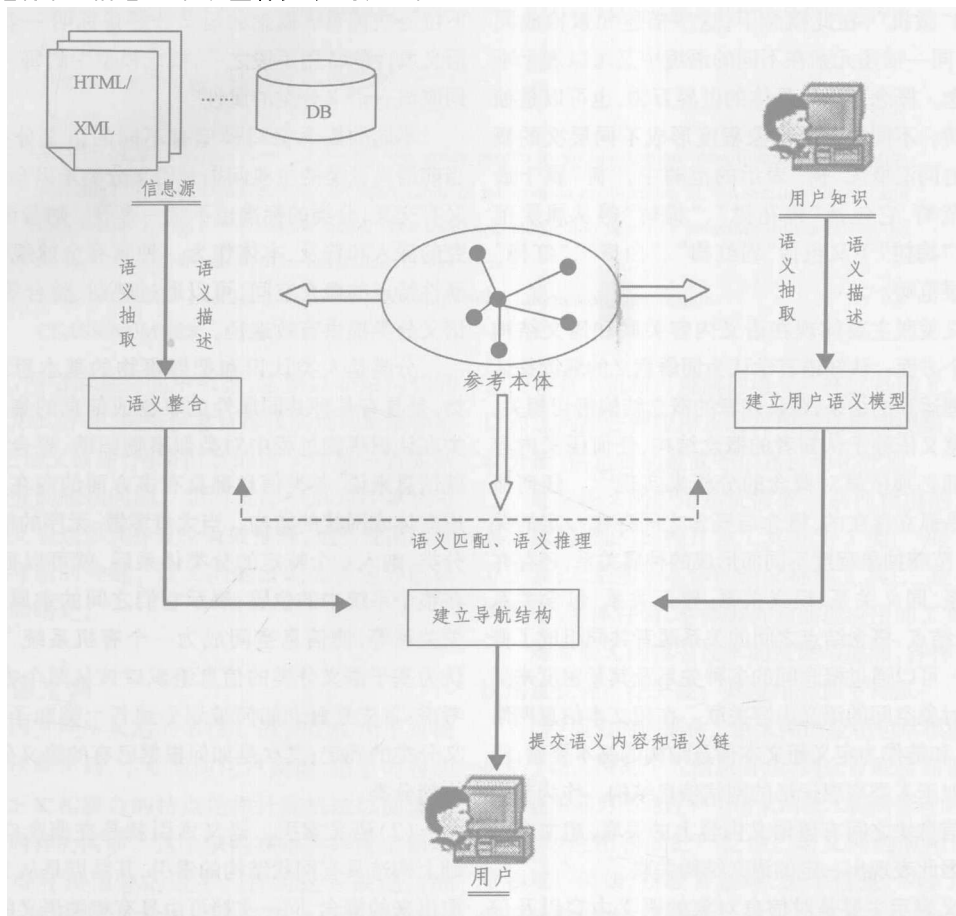


图2 语义导航模型框架

在语义导航模型中,可以通过提取领域知识的公共概念、属性、关系等构建参考本体,作为领域知识的整体视图,为信息源提供统一的语义。参考本体是语义整合、建立用户语义模型的基础,也是建立导航结构的核心。语义导航模型主要由3部分组成。

(1)语义整合。信息资源的组成非常丰富,既包括大量HTML页面,也含有XML、RDF等半结构化文档,还包括拥有海量信息的数据库等。采用语义网相关技术,对文档进行语义抽取、标注与描述,将信息源中各种格式的数据转换为通用数据模型,然后建立各个数据模型和参考本体之间的映射关系。

(2)建立用户语义模型。用户语义模型是用户描述某一对象时所用词汇及其关系集合的语义标注,反映用户对某一对象的理解,是用户的概念模型。不同知识背景的用户对信息的理解各不相同。网站导航系统设计中,如果不考虑用户对信息理解的差异,将在很大程度上影响导航的效果。因此,构建用户语义模型是导航的关键。用户语义模型的构建主要有两个方面:首先是按照一定规则确定用户的概念集合,即在分析用户兴趣爱好、信息需求、浏览行为等基础上,抽取出概念集合;其次是将用户的概念集合与参考本体进行语义映射,确定和用户概念相匹配的概

念以及概念之间的关系,建立用户的语义空间。通过将用户概念空间与参考本体进行语义映射,可以建立信息源和用户之间统一的语义模型。用户语义模型是各种导航方法的基础。

(3)建立导航结构。在参考本体的支持下,采用一定的导航方法(如基于用户需求的个性化导航、基于语义结构的浏览导航等),根据用户语义模型,通过语义匹配、语义推理等,查找具有相关语义内容的文档,在此基础上将满足用户需求的相关信息进行合

理组织,以一定的方式呈现给用户。

4 基于语义网的智能导航实例分析

目前国外已经开始基于语义网的导航研究,并建立了一些比较成功的试验系统,如 Magpie 系统、COHSE 系统、Onto Portal 系统、ESKIMO 系统等。下面将以 Magpie 系统为例,探讨语义网技术在智能导航系统中的实际应用。

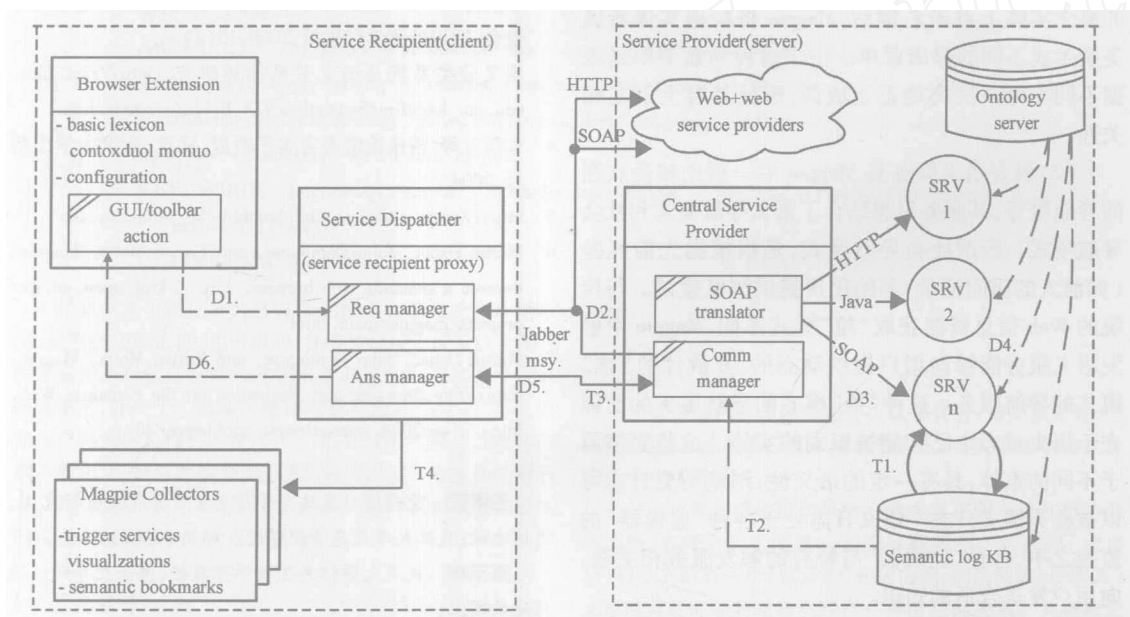


图3 Magpie 的整体框架结构

Magpie 导航系统是英国开放大学(The Open University)KMI 学院的一个研究项目。Magpie 起初虽然只是语义网上的浏览器插件,但随着进一步扩展和完善,现在已经集成了语义查询和其他语义服务,能提供高层次的智能导航。

4.1 Magpie 的体系结构

为了提供语义导航服务, Magpie 的技术组成主要可以分为两部分: Service Provider 组件和 Service Recipient 组件,如图 3^[6]。Magpie 的 Client 端的组件主要包括:

(1)Magpie 浏览器插件。它控制用户与 Magpie 之间的整个交互过程。Magpie 主要包含一系列用户界面组件,其主要任务是将 Web 页面上可识别的实体进行可视化处理和基于浏览过程生成上下文相关的导航菜单等。

(2)Magpie 服务调度器。这是一个专门的语义路由器。主要任务是管理 Magpie 浏览器插件和 Service Provider 端的合理通讯。Magpie 服务调度器将用户需求以及 Service Provider 端的需求答复按 XML 格式进行统一编码,并传递给对方。

(3)Magpie 收集器。它主要用来提供触发语义服务和书签导航。

4.2 Magpie 导航分析

Magpie 可以看成是一套嵌入在 Internet Explorer 浏览器中,以辅助用户浏览 Web 信息的工具。与其他标注工具不同的是, Magpie 并不是依赖手工进行 Web 页面标注,而是利用本体并根据用户的需求生成不同的语义层。用户选择了感兴趣的本体后, Magpie 可以根据用户选择的本体,自动与 Web 页面上的语义层进行交互。用户可以对识别出来的实体随机

指定一定的语义服务项目,以达到基于语义的 Web 浏览。Magpie 还可以根据语义浏览日志提供社会导航。可以说, Magpie 所提供的功能都是基于本体推理的。下面分析其主要的导航功能。

(1) 基于用户需求的智能导航。其目的是为了向用户提供上下文相关的知识,即随着用户浏览意图的不同,提供不同的信息,实现动态导航。在 Magpie 中,不同的本体拥有不同的类,不同的类别又拥有不同的属性。浏览 Web 页面时,用户在属于不同类别的概念实体上点击右键后, Magpie 将根据实体所属类别生成不同的弹出菜单。由于这种导航菜单是根据不同本体及其类动态生成的,因此具有上下文相关性。

(2) 触发语义服务是 Magpie 中一项比较受欢迎的导航服务,其服务思想结合了语义导航模式和社会导航模式。所谓社会导航模式,是指根据先前经验(如前人的访问记录)等给用户浏览建议。与传统的 Web 信息资源获取“拉”模式不同, Magpie 中触发语义服务能够向用户提供动态的、开放性的“推”模式的导航服务。这种“推”模式的导航服务的前提在于历史日志中已存储被识别的实体。这些实体属于不同的本体,具有一定的语义性,浏览历史日志可以被称为语义日志。语义日志处于各种“监视器”的监控之中,各种“监视器”与特有的触发服务相关联,向用户发送合适的知识。

(3) 书签导航。Magpie 通过使用语义日志能够自动跟踪用户浏览过程中所识别出的概念,并且把这一系列概念作为用户的浏览历史存放在语义日志知识库中。由于 Magpie 采用基于本体的推理,语义日志中所记录的用户浏览历史不再单纯是用户线性的

浏览行为,而是基于语义的浏览记录。用户在查阅其浏览历史记录时,得到的不再是一堆网页地址,而是一系列具有明确语义的实体所构成的概念空间^[7]。

参考文献

- 1 Berners-Lee, T., J. Hendler, and O. Lassila. The Semantic Web. Scientific American, May 2001. <http://www.scientificamerican.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21&catID=2>
- 2 郎天万, 蒋勇. 概念结构对语义原子论和语义场理论的整合. 四川外语学院学报, 2000, 16(2)
- 3 语义分类系统及语义关系描述研究. http://icl.pku.edu.cn/doubtfire/Semantics/973_Beida/contents.htm
- 4 张琪玉等. 情报检索语言实用教程. 武汉: 武汉大学出版社, 2004
- 5 http://www.intsci.ac.cn/ckmpbbh/ko_seminindex.jsp
- 6 Martin Dzbtor, John Domingue, and Enrico Motta. Magpie: towards a semantic web browser. <http://kmi.open.ac.uk/projects/magpie/main.html>
- 7 Martin Dzbtor, John Domingue, and Enrico Motta. Magpic: Supporting Browsing and Navigation on the Semantic Web. <http://iswc2004.semanticweb.org/demos/40>

王翠波 艾丹祥 武汉大学信息管理学院博士研究生。
通信地址: 武汉大学信息管理学院。邮编 430072。

张玉峰 武汉大学信息管理学院教授, 博士生导师。通信地址同上。

金燕 郑州大学信息管理系教师, 博士。通信地址: 郑州大学信息管理系。邮编 450001。

(来稿时间: 2005-07-05)

关于来稿中“参考文献”著录的两点说明

请向本刊投稿的作者, 除要严格按照国家标准《文后参考文献著录规则》对参考文献予以著录外, 还应注意以下事项:

1. 凡文后有参考文献的论文, 应在论文正文的相应位置, 用角注形式标出参考文献的序号。
2. 参考文献如果是网上文献, 著录项目不应省略责任者和文献题名; 并且不能只著录网站主页, 而应尽可能著录文献所在页面, 并加注作者从网上查得该文献的日期。

例如:

包冉. 国家网格在路上——2003 网格技术与应用研讨会侧记. <http://www2.ccw.com.cn/04/0402/b/0402b17-1.asp> (2003-03-11 查询)