

基于语义网的知识管理系统框架设计

张海霞, 吴江

(西北大学 计算机科学系, 陕西 西安 710069)

摘要:针对现有知识管理系统缺乏语义支撑的问题,结合知识的社会化、外在化、组合化、内在化螺旋动态转化过程,把知识的处理过程转化为知识俘获、知识存储、知识处理、知识使用这样一个动态循环过程,设计了一个基于语义网的知识管理系统框架,最后概述了开发系统用到的关键技术。

关键词:知识管理系统;语义网;知识地图;知识仓库;智能 Crawler

中图分类号:TP393.02

文献标识码:A

文章编号:1005-3751(2006)04-0046-03

Design for Knowledge Management System Framework Based on Semantic Web

ZHANG Hai-xia, WU Jiang

(Department of Computer Science, Northwest University, Xi'an 710069, China)

Abstract: Aiming at the problem about lacking semantics and combined the four processes of spiral knowledge translation: socialization, externalization, combination and internalization, knowledge process is translated into knowledge capturing, knowledge preserving, knowledge processing, using of knowledge. As a result, the framework of knowledge management system is presented, which is based on semantic Web. In the end, key technologies in developing the system are analyzed.

Key words: knowledge management system; semantic Web; knowledge map; knowledge storage; intelligent Crawler

0 引言

全球经济变得越来越知识化,现在知识已被广泛地认为是超越了土地、工具、劳动力而成为决定组织成败的最重要因素,作为知识经济的微观基础——企业,对此表现尤为明显。知识管理(KM)应运而生。知识管理是把存在于企业中的人力资源的不同方面和信息技术、市场分析乃至企业的经营战略等协调统一起来,共同为企业的发展服务。其出发点是将知识视为企业最重要的战略资源,把最大限度地掌握和利用知识作为提高企业竞争力的关键^[1]。

从信息管理的角度看,知识管理可以说是一种新的信息管理理论与方法,代表了信息管理的最新发展方向。知识管理系统(KMS)作为实现知识管理的计算机系统,是知识管理的实施平台。

目前,市场上虽然已出现了许多知识管理系统软件。但是,有关知识管理技术研究和应用中的许多关键技术问题还处于探索阶段。现有的知识管理系统主要存在如下的缺点:

1)搜索信息不准确:现有的基于关键字的检索可能查到只包含特定的术语而意思不同的无关信息,也可能错过与想要得到的内容使用的意思相同而术语不同的信息。传统的信息获得主要关注的是给定的查询和所存储的信息之间的关联,缺乏与信息相关的上下文。

2)提取信息智能性差:当前,从信息源中抽取出相关的信息还需要人的浏览和阅读。这是因为自动的智能代理(Agent)不具有从文本信息中抽取出相关信息的常识,也不能够整合分散在各个不同知识源中的信息。

3)难以维护:当资源增多时,维护弱结构的文本资源是既困难又费时的活动。保持这些资源的一致、正确和更新要求有助于发现不规则事物的语义形式表示。

4)自动文本产生缺乏语义:自动文本产生使按用户或其它恰当的方面自动重组恰当的站点成为可能,而来自于半结构化数据的半结构化信息表示的产生,需要这些信息源语义上机器可理解的表示法。

语义网是新一代的万维网,其目标是为万维网上的信息提供具有计算机可理解的语义,从而能够解决传统知识管理系统所存在的缺陷。文中主要介绍了一个基于语义网的知识管理系统框架。

1 语义网简介

语义网的概念是由万维网之父,英国人 Tim Berners

收稿日期:2005-07-01

基金项目:陕西省科技攻关项目(2003k05-g32);陕西省教育厅产业化培育项目(02JC47)

作者简介:张海霞(1979-),女,山西朔州人,硕士研究生,研究方向为语义网、知识管理。

-Lee 提出。现在万维网已成为人们获取知识的主要手段,但其在信息表达和检索方面存在缺陷,原因主要在于其设计目的是面向用户直接阅读与处理,而没有提供计算机可读的语义信息,因此限制了计算机在信息检索中的自动分析处理以及进一步智能化的信息处理能力。而语义网的目标是为万维网上的信息提供具有计算机可理解的语义,从而满足智能代理对万维网上异构、分布信息的有效检索和访问,实现网上信息资源在语义层上的全方位互联,并在此基础上,实现更高层的、基于知识的智能应用。

语义网采用多层次的表示框架 XML (eXtensible Markup Language), 并引入文档模式描述文档结构,但 XML 没有对结构本身的语义进行描述。如果不知 XML 元素的语义,计算机仍不能理解信息。需一个更高的层次来描述结构的语义,即 RDF (Resource Description Framework)。RDF 定义了元素之间的关系,表现为三元组集(类似于句子的主体、动词、客体)。XML 加上 RDF 相当于人工智能中的语义网络 (Semantic Network), 要进行简单推理使计算机相互理解,还需要有一套共同的标准概念体系,即本体 (Ontology)。XML + RDF + Ontology 构成了计算机相互理解的基础。在 Ontology 层之上还可以有更高的为各种 Ontology 进行通信提供支持的描述层次。语义网每加一个层次,计算机在知识处理上就多一份能力。可见语义网是发展知识管理很有前景的平台。

2 基于语义网的知识管理系统框架

知识管理系统中的主要过程是:发现与目前问题相关的可能知识和提供能用来分解一些问题的可能知识。从可编码性角度出发知识可分为隐性知识和显性知识,显性知识又包括形式化的专门规则和(多媒体)文档。为了更有效地搜索第二类显性知识,可以使用一些基于本体的声明(采用 If-Then 的形式)来索引文档的内容。

图1勾画了文中提议的基于语义网的知识管理框架,它反映了在分布式环境中知识转换的多样性:知识能从各种来源收集到和能以不同的形式收集,然后以相同的表示形式来存储、处理。该框架是基于现存的语义网工具 OntoMat, OntoMat - CRAWL, Ontobroker (http://ontobroker.semanticweb.org, http://ontobroker.semanticweb.org)等,为了操作更多的数据形式,对它们应该做轻微的扩展。

文中所提出的知识管理框架包含下面的过程:知识俘获;知识存储;知识处理;知识使用。所有过程都是以某种方法与领域本体相关的。下面对各个部分别进行描述:

1)文中定义了4种知识来源:专家知识、遗留知识库、元数据、文档。个人拥有的目前还无法形式化的隐性知识看成是专家知识,是推动创新的主要动力;遗留知识库是现有知识库和外部知识库的集合,是非常有价值的可共享知识资源,在解决问题时可以作为参考;在网络上扩展了的元数据应该是未来可分享知识主要的知识源,智能

Crawler 有收集元数据的能力;文档中的知识是非形式表示的,但文档的内容能够被基于本体的索引正式地规定。潜在的过程被称为语义注解,它可以由 OntoMat 注解框架来支持。通过使用信息抽取技术,使注解具有半自动性是很有可能性的。

2)智能 Crawler 是能够从网上自动获取相关信息并进行形式转化的工具。

3)转换器可以把遗留知识库中的知识转化成知识仓库中能够存放的形式化知识。

4)知识仓库是一个用某种方法组织起来的关系数据库,它能有效地存贮和访问元数据。

5)知识处理组件能够更加有效地处理被存储的知识,特别是对于以规则形式表示的知识处理。

6)知识地图是知识目录和领域专家的导航,它允许对所描述的组织知识资源进行处理、浏览和形象化。知识地图是组织协作工作的重要组成部分,它可以作为电子邮件的附录发送到组织的所有成员信箱中,从而使组织的所有成员能共享这种明确清晰的知识。

7)知识的使用与用语义网激活的知识入口相关^[2]。这种方法的主要好处就是使用了条件声明来对知识源进行语义注解。以这种方法把注解中使用的声明放在彼此的上下文中,从而有效地引导搜索知识。

8)知识推送能根据自身的刷新时间表发布信息。用户可以预订推送信息服务,也可由系统试探能让用户感兴趣的其它信息,这时用户得到的可能是全屏报道,也可能是屏幕底部显示的大字标题,用户可以选择需要查看的信息流,再由系统精选发送给用户。

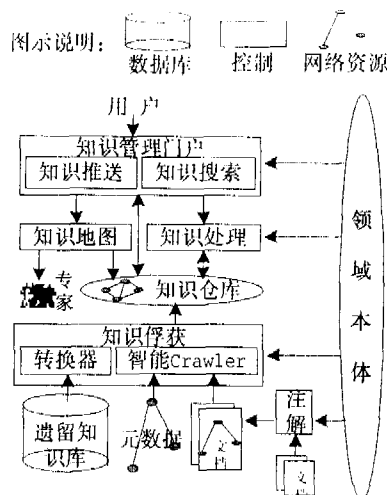


图1 知识管理系统框架

3 系统开发关键技术

利用知识管理系统来辅助知识管理,受许多因素的影响,下面简单介绍一下系统中的关键技术。

3.1 知识地图

知识地图技术是用于帮助人们知道在什么地方能找到知识的管理技术,其实质是利用现代化信息技术制

作的组织知识资源的总目录及各知识项目之间关系的综合体,是组织中协作工作的重要组成部分^[3]。知识地图不仅能清楚地揭示组织内部或外部相关知识资源的类型、特征及知识之间的相互关系,还能揭示组织的知识结构、业务流程、员工激励制度、客户承诺以及组织用以创造和利用知识的技术。知识地图所描绘的对象主要有人、显性(或编码化)的知识以及过程(或方法)。这 3 种对象通常所描述或指向的知识分别为隐性知识、显性知识和隐藏知识。在知识地图中一般使用抽象的符号或图像来表示这些对象。

3.2 知识仓库

知识仓库包含各种各样的知识,有外部知识(如:通用的方法、基本理论知识);结构化的内部知识(如:调查报告,面向产品的市场销售材料和方法以及各类文档);非正式的内部知识(如:各种技术诀窍、经验教训以及各类讨论及在线会议等所组成的数据库)等,而且这些知识以多种方式存在,如 Office 文档、Web 页面、数据库形式等,能否有效地整理、归类、进而分析和挖掘其内在的隐含知识,使知识仓库更加有序条理化,知识范围更加宽广,从而更好地被检索和利用,是整个知识管理系统是否有效的基础^[4]。

文中提出了一种基于本体的知识仓库。知识仓库系统中可以存储许多文档,这些文档可以列出其相应的关键字和摘要,通过本体的协助,结合文档的内容,可以判断该文档的专业范畴,以此对文档分类。一个专业本体给出了该专业中各个术语的定义和术语之间的关系,通过关键字,查询本体中的相应术语,可以找出与该术语相关的信息,依据此信息可以在知识仓库中检索出用户所需求的信息。当用户提出查询请求的时候,智能代理与用户交互,根据用户提出的关键字,查询整个本体库,判断该关键字的专业领域,并且将其中的相关术语提交给用户。

3.3 智能 Crawler

智能 Crawler^[5]的主要任务是自动发现主题下新的相关资源,并且加入到集合中。一般的 Crawler 只是简单地进行深渡优先或广度优先的搜索,不具智能性,不会根据用户反馈的信息有选择地快速扩展。文中所提出的智能 Crawler 能根据文本声明来判断文档是否值得扩展。提高了效率,还通过计算新文档与原集合标准向量的匹配程度,有选择地加入新资源。

智能 Crawler 的工作流程如下:

(1) 新建集合 A, 把原集 R 中文档加到 A。

(2) 若 A 为空, 停止扩展, 否则解析集合 R 中文档所包含的超链接 URL, 对不在集合 R 中的 URL, 满足条件 1(见下文)的组成一个新的集合 T。

(3) 若 T 为空, 停止扩展, 否则对集合 T 中的所有 URL 对应的文档, 满足条件 2(见下文)的加入到原来集合

R 中, 成为 R', 使 $A = R' - R$, $R = R'$, 清空集合 T。

(4) 判断是否满足停止条件, 如果满足, 停止扩展, 否则, 跳至 1。

条件 1 文本声明对主题的描述精度大小。考虑链接 $p \rightarrow q$, p 中有若干形似如下的链接标记:

文本 1 $\langle a \ href = "http://..." \rangle$ 文本声明 $\langle /a \rangle$ 文本 2

其中文本声明对目标 URL = http://... 有比较精确的描述, 文本 1 和文本 2 次之。实验表明, 文本 1 和文本 2 的长度取 50 字节比较合适, 离文件声明远的文字精度急剧降低, 可以不用考虑。文件声明和文本 1, 2 总权重按如下公式计算:

$$W = \sum_i W_i$$

知识管理系统设定一个阈值 c, 当 W 大于 c 时, 此目标 URL = http://... 加入到集合 T, 否则不加入。经过条件 1 的过滤, 可以减少很多无关的文档, 提高了扩展速度和精度。

条件 2 新文档的向量和主题标准向量的匹配程度。对于满足条件 1 的 URL, 知识管理系统会从 Web 上取得它的全文, 得到文档向量 V_i , 计算向量 V_i 和标准向量 V 的距离 $D(V_i, V)$, D 小于设定的阈值 d 的, 可以加入到集合中做进一步的扩展, 否则, 不加入。经过条件 2 的过滤, 进一步提高了结果精度, 但实时地分析每个文档的向量也降低了效率。

4 结束语

语义网是发展知识管理非常有前景的平台。文中介绍了一个在语义网上实现知识管理系统的框架, 为更好地实现知识管理提供了一种新的可能性。它的主要优点是能够克服现有知识管理系统在信息搜索方面所存在的缺陷。

参考文献:

- [1] 王德禄. 知识管理的 IT 实现——朴素的知识管理[M]. 北京: 电子工业出版社, 2003.
- [2] Stojanovic N, Maedche A, Staab S, et al. SEAL - A Framework for Developing Semantic PortALS[R]. Vancouver: ACM, 2001.
- [3] Lubell J. Structured Markup on the Web: A Tale of Two Sites[M]. Boston: Massachusetts Institute of Technology Press, 1999.
- [4] OFFSEY S. Knowledge management: linking people to knowledge for bottom line results[J]. Journal of Knowledge Management, 1997(2): 117 - 119.
- [5] Chakrabarti S, van den Berg M, Dom B. Focused Crawling: A New Approach Topic - specific Web Resource Discovery[M]. Holand: Elsevier Science B. V., 1999. 545 - 562.