

Web 2.0 下基于案例的知识管理系统

李 锋

(华南理工大学工商管理学院, 广州 510640)

摘 要: 为实现 Web 2.0 环境下网络知识的获取和共享, 提出一种基于案例的知识管理系统。采用案例推理的方法实现知识管理, 根据 Web 2.0 网络环境的特点, 在案例表达阶段, 为每个案例设置标签属性, 使其能实现开放性分类。在案例检索阶段, 将领域本体引入相似度的计算。利用神经网络算法、用户录入案例标签维护案例库和本体库。实验结果表明, 该系统能快速检索出具有较高相似度的历史案例。

关键词: 知识获取; 知识共享; 标签; 案例检索

Case-based Knowledge Management System Under Web 2.0

LI Feng

(School of Business Administration, South China University of Technology, Guangzhou 510640, China)

【Abstract】 In order to realize knowledge acquisition and knowledge sharing in Web 2.0, a case-based knowledge management system is presented. According to bottom-up approach of knowledge innovation in Web 2.0, methodology of Case-based Reasoning(CBR) is adapted to manage knowledge. In phase of case representation, tags are introduced to label each case for better case classification and identification by system users. In phase of case retrieval, domain ontology is also introduced to find the most similar history case. To maintain case base and ontology base, artificial neural network algorithm and tags created by system users are used. Experimental results show the system can improve the efficiency of the retrieval precision, and the history case has high similarity.

【Key words】 knowledge acquisition; knowledge sharing; tag; case retrieval

DOI: 10.3969/j.issn.1000-3428.2012.02.027

1 概述

Web 2.0 是在传统互联网应用(Web 1.0)基础上发展而来的一类崭新的互联网应用统称。具有代表性的 Web 2.0 应用包括即时信息(Instant Messaging, IM)、博客/微博(Weblog/Microblog)、社会网络(Social Network)、百科全书(Wiki)等。相比传统互联网应用, Web 2.0 强调互联网只是一个信息的公共载体/平台, 每一位网络用户既是网络信息内容的消费者, 同时也是网络信息内容的创造者。

因此, 随着 Web 2.0 技术的发展和推广, 如何从分散的、异构的网络信息源中快速地获取信息和知识, 并实现知识的共享、重用和创新等管理问题已经成为当前的研究热点问题。基于上述思考, 本文提出一种 Web 2.0 下基于案例的知识管理系统。

2 研究现状

Web 2.0 的发展造成信息不再是由少数人集中控制, 而是由所有网络用户共同作用的结果。例如, 网络用户通过各种网络平台的交流, 形成新型的用户群体-网络社群。在社群中, 成员通过个人思维活动, 并伴随着成员之间的信息交流和共享, 不断地创造着新的知识。并且, 随着 Web 2.0 技术的成熟, 网络社群所创造出的知识已经成为一类重要的创新知识的源泉^[1]。

但是, 网络社群中松散的成员具有开放性、自主性等特点, 这使得其群体创造出的知识具有鲜明的自下而上的特征。因此, 研究如何从网络社群中获取知识具有十分重要的研究意义。鉴于 Web 2.0 环境下知识自下而上的形成过程, 本文

采用知识工程中的案例推理(Case-based Reasoning, CBR)方法建立其系统概念模型, 并在此基础上实现知识的管理。由于案例推理方法不仅是一种推理技术, 而且是一种系统分析的方法论^[2], 它将历史已经解决的问题, 及其解决的方法存放在案例库中。

在求解新问题时, 首先从案例库中检索出最接近的历史案例/问题, 通过借鉴并调整其解决方法, 从而解决新问题。在案例库中, 案例作为历史经验知识的一种载体, 同样具有知识共享、知识重用、知识创新等职能^[3]。

案例库以及案例库的知识管理能够适用于网络中的知识管理, 并且以案例的形式管理知识, 能够避免传统知识系统中知识获取的瓶颈问题, 特别适用于 Web 2.0 环境下由网络用户自行建立(自下而上)的知识库。

由于传统的案例推理方法不能简单套用于 Web 2.0 环境下网络知识管理, 因此本文主要研究适合 Web 2.0 环境下网络知识案例管理的相关技术和方法。

3 网络知识的案例管理系统

3.1 系统架构

案例管理系统架构如图 1 所示。

基金项目: 华南理工大学研究生重点课程建设基金资助项目(2000 8028); 华南理工大学中央高校基本科研业务费专项基金资助项目(2009zm0079)

作者简介: 李 锋(1975—), 男, 副教授、博士, 主研方向: 智能信息处理, 复杂系统建模

收稿日期: 2011-07-19 **E-mail:** fenglee@scut.edu.cn

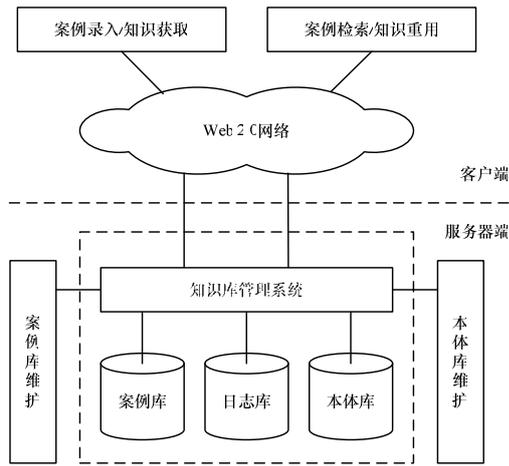


图1 案例管理系统架构

如图1所示,本文系统采用客户机/服务器架构:在客户端,网络用户通过浏览器完成案例录入和案例检索。同时,用户的案例录入和案例检索活动都将记录在服务器端的日志库中。在服务器端,系统管理员通过知识库管理系统系统界面维护后台的2个知识库,即案例库和本体(Ontology)库,以案例的形式实现网络知识的知识表达,并以案例库的形式存储网络知识,其关键在于结合案例的表达形式,选择最适当的案例检索方法,从而实现方便、快捷的知识重用。除此之外,为了提高案例检索的检索效率和检索精度,还必须对已经存放在案例库中的历史案例进行维护(全生命周期、在线式的维护),即通过修订、合并,甚至删除历史案例,以及修改案例的表达形式、索引机制等案例库维护操作以减少案例库中的冗余案例和信息。

3.2 案例的表达

在本文系统中,案例的表达采用基于框架(Frame)的知识表达。在知识表达的框架理论中,一个框架通常由若干个槽(Slot)的结构组成,每个槽又可根据实际情况拥有若干个侧面(Facet),每个侧面也可以拥有若干个侧面值(Value)。

一个用于描述计算机建模练习的案例可以用框的形式描述如下:

- 计算机建模练习案例 No.XX
- 槽 1: 模型描述
 - 侧面 1: 输入过程
 - 侧面值: 实体到来间隔时间分布; ……
 - 侧面 2: 服务机构
 - 侧面值: 服务时间分布; 服务规则; ……
 - ……
- 槽 2: 模型分析
 - 侧面 1: 计算机仿真模型错误原因
 - ……
- 槽 3: 模型修正
 - 侧面 1: 输入过程
 - 侧面值: 实体到来间隔时间分布; ……
 - 侧面 2: 服务机构
 - 侧面值: 服务时间分布; 服务规则; ……
 - ……

在传统案例表达方式的基础上,结合 Web 2.0 网络特点,为每个案例设置标签(Tag)属性,用于由网络用户自定义该案例在案例检索阶段使用时的一个或多个的检索关键词。

通过标签能够对每个案例实现开放性分类,对于每个案

例可以有多个标签,这样的话,可以将具有相同标签的内容联系起来。在检索过程中,一方面可以给出该标签下的相关的案例,另一方面还可以给出与其相关的标签。

3.3 案例的检索

案例检索算法采用基于语义的相似度计算算法^[4]。通过引入领域本体,能够更加准确地理解由用户设置的案例标签,以及标签之间的相关性。

由于案例的属性值包括定量指标集合(传统案例表达部分)和定性描述指标集合(Web 2.0 文本标签部分)2个部分,因此定义问题案例 C_x 与历史案例 C_y 之间相似度为:

$$Sim(C_x, C_y) = \frac{(\omega_a \cdot S_a(C_x, C_y) + \omega_b \cdot S_b(C_x, C_y))}{(\omega_a + \omega_b)} \quad (1)$$

其中, $S_a(\cdot)$ 、 $S_b(\cdot)$ 分别代表案例 C_x 、 C_y 定量指标集合之间的相似度和定性指标集合之间的相似度; ω_a 、 ω_b 代表 $S_a(\cdot)$ 、 $S_b(\cdot)$ 的权重。

案例 C_x 、 C_y 的定量指标集合之间,相似度计算式 $S_a(\cdot)$ 采用最近相邻算法,即^[5]:

$$S_a(C_x, C_y) = \frac{\sum_{i=1}^n (\omega_i \cdot (1 - d_i(f_i^x, f_i^y)))}{\sum_{i=1}^n \omega_i} \quad (2)$$

其中, $d_i(\cdot)$ 表示标准化处理后案例 C_x 、 C_y 第 i 个属性值之间的距离; ω_i 表示案例第 i 个属性的权重。

案例 C_x 、 C_y 定性指标集合之间的相似度计算式 $S_b(\cdot)$ 采用基于语义的案例相似度计算式,即:

$$S_b(C_x, C_y) = \begin{cases} r(t_x, t_y) & t_x \in O, t_y \in O \\ c(t_x, t_y) & t_x \notin O, t_y \in O \end{cases} \quad (3)$$

如式(3)所示,当问题案例 C_x 的标签 t_x 在系统本体库 O 中已经定义,则标签之间的相似度由语义库中标签之间的关系链定义^[4];否则,标签之间的相似度由计算字符串之间相似度的 Dice Coefficient 方法定义,即:

$$c(t_x, t_y) = 2 \cdot n_i / (n_x + n_y) \quad (4)$$

其中, n_x 、 n_y 分别表示标签 t_x 、 t_y 中字符的个数; n_i 表示标签 t_x 、 t_y 相同字符的个数。

3.4 知识库的维护

系统管理员需要定期对后台的知识库(包括案例库和本体库)进行维护,目的是从知识库中剔除冗余或者错误的知识,从而提高系统运行的性能和效率。

对于案例库的维护工作主要是从案例库中剔除冗余的历史案例,并且从表达案例的属性集合中剔除冗余的案例属性特征^[3,6]。本文应用人工神经网络(BP 网络)算法来维护案例库中案例的代表性^[3],并通过统计用户进行案例检索时案例属性利用率(日志库统计信息),对案例属性和案例属性的权重进行调整。

对于本体库则是根据用户录入的案例标签进行维护。在本体库中,定义并量化标签之间的语义关系^[4]。标签之间的关系包括“等同”关系、“类似”关系和“隶属”关系3种。由此量化标签之间相似度的方法如下:

- (1)建立只包含“等同”关系的标签的邻接矩阵 A 。
- (2)由矩阵 A 计算得到标签之间的可达矩阵 R_A 。
- (3)矩阵 R_A 中 1 元素不变; 0 元素修正为 1, 当其所代表的标签之间存在着“类似”关系。从而得到新的邻接矩阵 S 。

(4)由矩阵 S 计算得到新的可达矩阵 R_S 。

(5)矩阵 R_S 中 1 元素不变; 0 元素修正为 1, 当其所代表的标签之间存在着“隶属”关系。从而得到新的邻接矩阵 F 。

(6)由矩阵 F 计算得到新的可达矩阵 R_F 。

由矩阵 R_A, R_S, R_F 定义标签 t_x, t_y 之间的相似度为:

$$r(t_x, t_y) = \begin{cases} 1 & R_A(t_x, t_y) = 1 \\ 0.7 & R_S(t_x, t_y) = 1 \\ 0.5 & R_F(t_x, t_y) = 1 \\ 0 & R_F(t_x, t_y) = 0 \end{cases} \quad (5)$$

其中, $R_A(t_x, t_y)$ 、 $R_S(t_x, t_y)$ 、 $R_F(t_x, t_y)$ 分别表示标签 t_x 、 t_y 在矩阵 R_A 、 R_S 、 R_F 中的取值。

4 系统实现与测试

结合所承担的课题, 在 Web 2.0 平台上开发一套基于案例的知识管理原型系统, 用于管理“供应链系统仿真”课程练习题。

供学生操作的案例录入系统界面如图 2 所示。

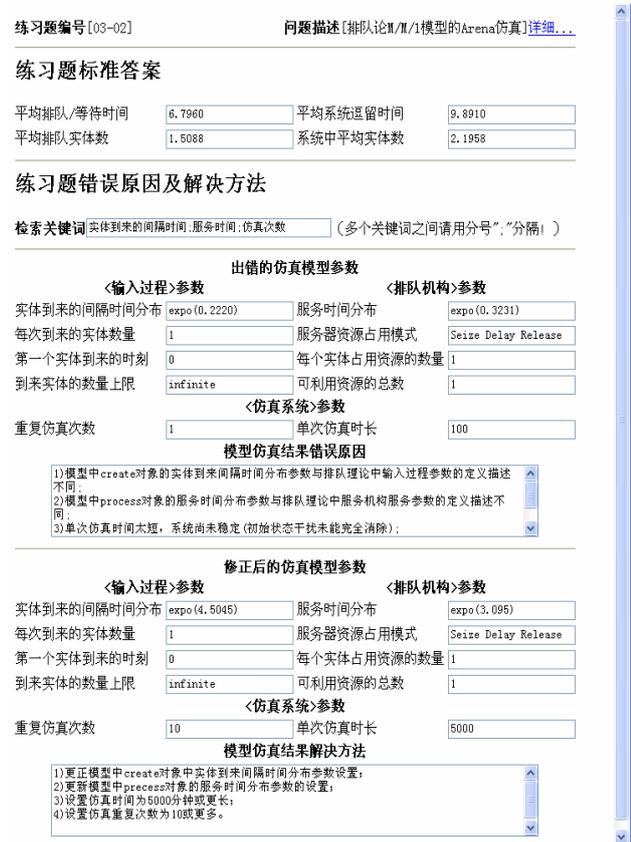


图 2 案例录入系统界面

由图 2 可知, 传统的案例检索算法难以计算相似度的文本属性(无结构文本属性的相似度计算因为涉及到自然语言理解等难点而难以计算和比较), 通过增加 Web 2.0 下特有的标签技术, 并结合领域本体信息从而得到了解决。

在系统试运行阶段, 建立起的标签间语义关系(部分)如图 3 所示。

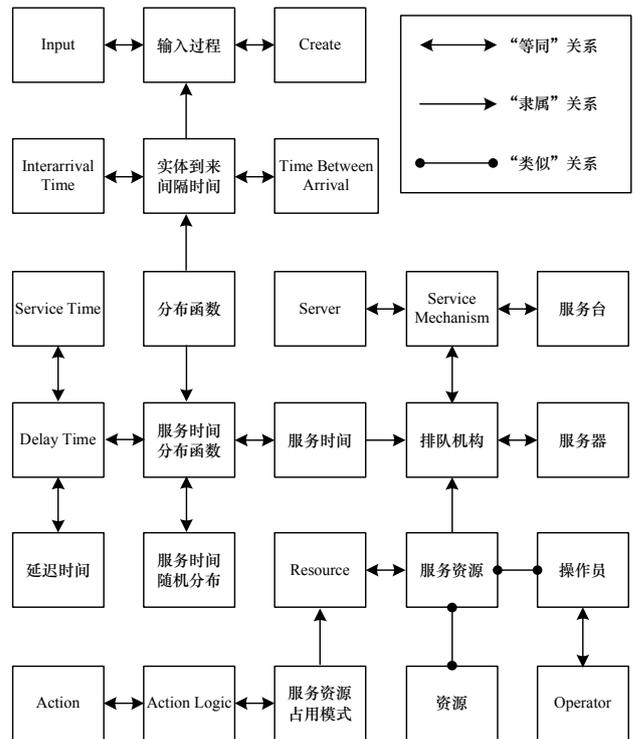


图 3 标签间语义关系

5 结束语

本文针对 Web 2.0 环境下知识形成过程中自底向上的特点, 提出基于案例的知识管理系统。在案例化的知识获取中, 通过引入 Web 2.0 特有的标签技术, 能够辅助并提高知识重用阶段的检索效率。在检索阶段, 通过引入本体实现基于语义相似的案例检索, 使其检索出的历史案例具有较高的相似度。通过对原型系统进行测试, 证明该系统有效。

在原型系统的初上线试运行过程中, 通过对后台日志库的监控, 发现标签设置、内容非常发散, 这导致在案例检索阶段得到的相似案例数量多且繁杂。因此, 最相似历史案例子集需要采取更加友好、系统的展示方式。这是下一阶段的研究重点。

参考文献

- [1] 邓胜利, 胡吉明. Web 2.0 环境下基于群体交互学习的知识创新研究[J]. 情报理论与实践, 2010, 33(2): 17-20.
- [2] Watson I. Case-based Reasoning is a Methodology not a Technology[J]. Knowledge-based Systems, 1999, 12(5): 303-308.
- [3] 李 锋, 冯 珊. 基于人工神经网络的案例检索与案例维护[J]. 系统工程与电子技术, 2004, 26(8): 1053-1056.
- [4] 李 锋, 魏 莹. 分布式环境下基于语义相似的案例检索[J]. 计算机工程, 2007, 33(9): 28-30.
- [5] 李 锋, 魏 莹. 基于偏好信息的案例检索算法[J]. 计算机工程, 2008, 34(24): 28-30.
- [6] 孙岩清, 尹树华, 王 技. 基于粗糙集的 CBR 系统属性约简改进算法[J]. 计算机工程, 2010, 36(10): 38-40.

编辑 刘 冰