

基于本体的面向短文本的饰品设计知识服务

郭 磊, 刘运通

(安阳师范学院 计算机与信息工程学院, 河南 安阳 455002)

摘 要: 针对产品设计过程中大量知识是以短文本的方式存储, 难以进行高效、准确的知识服务的问题, 提出了一种基于本体的短文本知识集成方法. 首先, 构建产品设计的领域本体; 然后, 根据本体所蕴含的语义信息, 对短文本进行分析, 从而找到蕴含用户所需知识的相关短文本, 并通过知识服务将其提供给设计人员. 并在某企业的饰品三维辅助设计环境下进行应用验证, 结果表明, 该方法能够提供高效、准确的饰品辅助设计知识.

关键词: 短文本; 本体; 语义; 知识服务; 知识获取; 知识转化

中图分类号: TP391

文献标识码: A

文章编号: 1000-7180(2012)08-0066-04

A Short-text-oriented Knowledge Service for Ornament Design Process Based on Ontology

GUO Lei, LIU Yun-tong

(School of Computer and Information Engineering, Anyang Normal University, Anyang 455002, China)

Abstract: In order to provide knowledge service for ornament design process more accurately and efficiently, a knowledge integration method based on ontology was proposed to acquire the knowledge implicated in short texts. As a precondition, a domain ontology should be constructed to express the semantic relations between the concepts in ornament design process, we could acquire the knowledge that were implicated in short texts through semantic analysis, and we could provide the knowledge to the designer by knowledge service. Finally, a knowledge service system for ornament design process in Rhinoceros was presented to verify the effectiveness of the method.

Key words: short texts; ontology; semantic; knowledge service; knowledge acquisition; knowledge transformation

1 引言

企业的产品设计过程中需要大量的知识, 并且这些知识大多蕴含在文本数据中, 首先需要对这些文本数据进行一定程度的自然语言处理, 获取其中所蕴含的知识, 才能进行知识服务. 文献[1]研究了基于模糊认知图和贝叶斯信念网络的产品设计知识获取; 文献[2]研究了基于人工神经网络的产品设计知识服务; 文献[3]研究了基于决策树和支持向量机的产品设计知识管理方法. 由于自然语言处理这个难题的存在, 这些方法取得的效果依然不够显著.

文献[4]研究了对专利文献进行文本挖掘, 对研发所需的专业知识进行管理; 文献[5]研究了通过聚

类分析来对短文本进行知识分类.

这些研究忽视文本中的语义信息, 难以提供高效、准确的知识服务.

文献[6]研究了基于领域本体的文本知识获取; 文献[7]研究了基于领域本体网络模型的中医领域文本知识获取. 但这些研究只是利用语义信息来解决文本数据知识管理中的部分难点问题, 还没有形成有广泛影响的基于语义分析的文本数据知识管理模型.

因此, 本文针对面向特定领域的短文本的知识服务问题, 提出了一种语义分析方法, 基于该方法, 在某企业的饰品三维辅助设计环境下进行应用验证, 取得了较好的效果.

收稿日期: 2012-01-16; **修回日期:** 2012-03-01

基金项目: 国家“八六三”高技术研究发展计划资助项目(2006AA04Z157, 2007AA040607)

2 知识服务的系统结构

系统包括五部分:数据源、知识库、数据分析模块、知识检索模块、用户接口模块,如图1所示。

数据源:需要进行处理的源数据,可以是文本文件中的数据,也可以是可能是数据库中的某个字段(该字段是一个短文本,这种情况也很普遍)。

知识库:短文本数据经过处理后,以一种特定的方式保存起来所形成的知识库。

知识库构建模块:根据领域本体的语义关系,解析文本文件中的语句,将短文本转化为知识;

知识检索模块:根据用户的输入,查找知识库中的相关知识,并将检索到的结果返回给用户。

用户接口模块:用户输入和输出的界面。

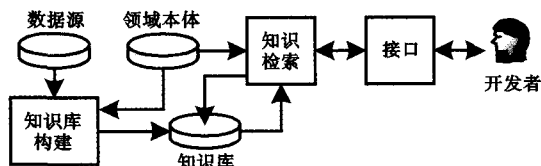


图1 知识服务的系统结构

3 构建知识库的方法

3.1 将短文本转化为知识的规则

从本质上来说,一个短文本就是一个短语,类似于依存语法中的依存关系^[8],在一个短文本(用 C_s 表示)中,任意实词 W_i (除去谓语中心词)均语义修饰于另外一个实词 W_{Gi} ,称 W_{Gi} 是 W_i 语义修饰目标。一个短文本可能具有多种语法分析方案;假设在某种特定语法分析方案 A_i 的情况下, V 为谓语中心词, S 为 V 的施动者, O 为 V 的承受者,如图2所示。

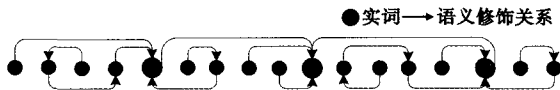


图2 短文本分析示意图

则将短文本 C_s 转化为知识的规则为:

$$C_s \rightarrow V(S, O) \quad (1)$$

$V(S, O)$ 作为一则知识,表示主体 S 针对客体 O 是执行了 V 操作。

知识转化的本质就是找到短文本的最佳主题词 S, V, O 。

3.2 知识转化模型

一个短文本 C_s 可能具有多种语法分析方案如

图3所示。图2仅仅是其中的一种特定语法分析方案 A_i 。假设 W_i 是语句中的一个实词且! ($W_i \in S, V, O$), W_{Gi} 是 W_i 的语义修饰目标,用函数 $\text{match}(W_i, W_{Gi})$ 表示在领域本体中两个概念的语义关系的值;那么,整个语句在语法分析方案 A_i 的情况下的语义关系 Value_{A_i} 可以用公式来表示:

$$\text{Value}_{A_i} = K_{SVO} * (\text{match}(S, V) + \text{match}(O, V)) + K_w \sum_{i=1}^n \text{match}(W_i, W_{Gi}) \quad (2)$$

K_{SVO} 是关于 S, V, O 语义关系值的权重系数, K_w 是普通词汇之间的语义关系值权重系数。

3.3 最佳语法分析方案的选取

根据式(2),针对所有可能的语法分析方案(如图3所示),通过计算获得到的整个语句的语义匹配值,最后根据语义匹配值计算结果,选择出具有最佳语义匹配值的语法分析方案,并将短文本转化为用式(1)表示的知识。

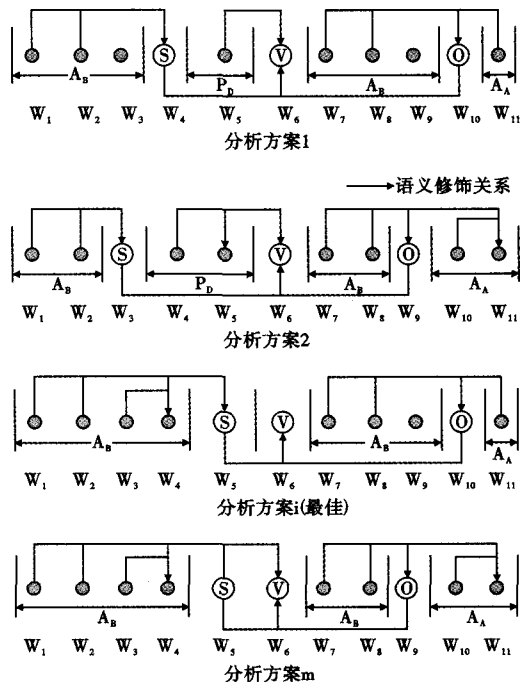


图3 知识转化的模型示意图

3.4 知识转化模型的求解方法

知识转化模型的求解思路:根据3.2中的模型,生成所有可能的语法分析方案,如图3所示,并根据式(2)计算出每种语法分析方案的语义关系值,选择出具有最佳语义关系值的语法分析方案作为结果,并使用式(1)将其转化一则知识。

在理论上求取最佳的语法分析方案,需要考虑极为复杂的语法规则,其分析方法也不成熟,但由于

以下原因,可以通过 3.3.1 节中的方法求取其近似解:

(1)蕴含产品设计知识的文本数据往往很短,可以不用进行复杂的语法分析;

(2)短文本中的实词大多属于专业程度很高的词汇,并且包含在领域本体中,可以进行语义计算。

3.4.1 近似求解方法

因此,在求取知识转化模型的近似解时,我们忽略了复杂的语法规则分析,仅仅根据语义关系,对语句进行分析;将语句划分为两个层次:简单句、复杂句(包含简单句的语句),并设计出两个层次的通用文法,其设计思路如图 4 所示。

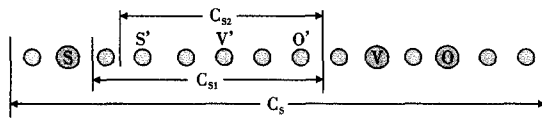


图 4 短文本的文法设计

(1)简单句的文法 G_1 的规则(图 4 中的 C_{s1} 、 C_{s2}):

$C_s \rightarrow SVO$; 2) $S \rightarrow WS$ | WS , $V \rightarrow WV$ | WV , $O \rightarrow WO$ | WO 。

在上述文法 G_1 中,假设在领域本体中,短文本中的词汇 $W1$ 、 $W2$ 之间具有 $W3$ 的语义关系,则可将 $W1$ 、 $W2$ 设置为 S 、 O , $W3$ 设置为 V 。

(2)在 G_1 中添加规则 $W \rightarrow C_s$,就形成了复杂句的文法 G_2 (如图 4 中的 C_s)。

算法 1(可归结简单子句的自底向上归结法)

步骤 1(找到所有的简单子句):针对 C_s ,进行文法 G_1 的 CYK 算法,找出所有可能的简单子句;

步骤 2(选取每个实词的最佳语义修饰目标):针对每个实词,穷举出所有可能语义修饰目标,选取具有最佳语义关系的实词作为语义修饰目标;

步骤 3 计算简单子句的最佳语义关系值;

步骤 4(归结):用 $W \rightarrow C_s$ 归结该简单子句;

步骤 5(选取最佳结果):递归归结,具有最佳全句语义关系值的归结过程所对应的语法分析方案是最佳语法分析方案;

步骤 6(知识转化):用规则 $C_s \rightarrow V(S, O)$ 将短文本转化知识。

3.4.2 语义关系值的计算

一般情况下,词汇 $W1$ 、 $W2$ 之间的语义关系值等于两个词汇在领域本体中的最短距离,类似于在 wordnet 与 hownet 中的计算方法。

3.4.3 知识转化实例

短文本数据的知识转化方法是知识服务系统的

关键,现以如下的实例来说明,假设短文本为:

- (1)老人适合佩戴树脂饰品;
- (2)水辅不能镀锌;
- (3)德国人厌恶绿色饰品。

根据本文的方法,对这些短文本进行知识转化,具体结果如表 1 所示。

表 1 知识转化情况表

短文本	知识表示(相当于谓词逻辑命题)
老人适合佩戴树脂饰品	佩戴(老人,树脂)
水辅不能镀锌	镀(水辅,锌)
德国人厌恶绿色饰品	厌恶(德国人,绿色饰品)

4 知识检索

当知识库构建完成后,即可从知识库内,对用户的提问进行知识检索。

知识检索时,要求用户提供 S 、 V 、 O 的关键词,然后从知识库中找到匹配的知识即可,并可以按照语义距离对知识进行排序。

假设用户提供的关键词是 S 、 V 、 O ,则知识 X (Y, Z)是所需的知识的条件如下:

$$(X \propto V) \cap (Y \propto S) \cap (Z \propto O) \quad (3)$$

式中, $X \propto V$ 表示概念 X 是概念 V 的子类或实例,如图 5 所示。

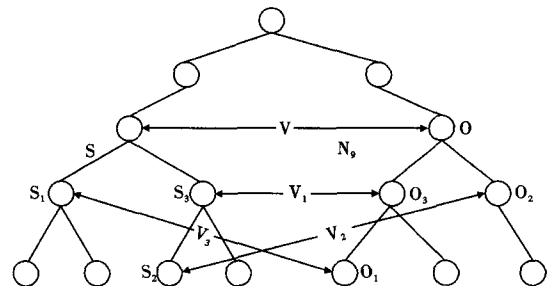


图 5 知识检索原理示意图

5 应用验证

某饰品公司采用三维饰品设计建模软件 Rhinoceros 4.0 SR3 进行产品辅助设计;并且具有大量的饰品辅助设计知识帮助文档,这些饰品设计知识大多是以短文本的方式存在的,现需要依据这些数据进行提供高效的知識服务。

5.1 系统实现

利用本文的方法,基于 Microsoft Visual Studio 2005 平台,利用 Rhinoceros SDK,开发了饰品设计知识服务系统,该系统以插件的形式嵌入到 Rhi-

noceros 中(如图 6 所示). 在该知识服务插件中,利用本文的方法,可以将短文本转化为知识,并检索出用户需要的相关知识.

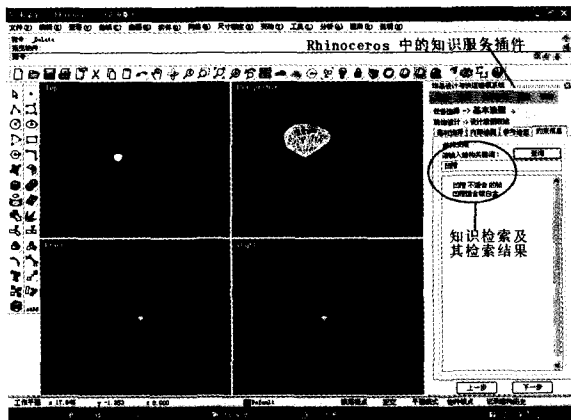


图 6 饰品设计知识服务界面

5.2 运行效果分析

在实验时,首先通过 protégé 构建一个包含 2 390 个概念的饰品设计领域本体;然后针对 3 450 条饰品设计短文本数据进行知识转化;用户就可以进行饰品设计知识查询,结果如表 2 所示.

表 2 实验结果

项目	结果/%	项目	结果/%
正确率	96	召回率	91

这里的正确率、召回率特指如下情况:根据用户所输入的关键词,检索出蕴含所相关知识的短文本的正确率与召回率.

6 结束语

由于很多产品设计知识蕴含在短文本数据中,针对这些语法结构比较简单短文本来说,由于其中的实词大多属于非常专业词汇,因此可以通过本文的知识转化语义模型进行处理,将其转化为特定的知识表示形式,从而就可以进行高效的基于语义的知识查询和知识服务了,并在饰品设计系统中加以应用,取得了较好的效果.

参考文献:

- [1] Kim Yun Seon, Kim Kyoung-Yun Joseph, Cheah Wooi Ping, et al. Causal design knowledge acquisition by constructing BBN through FCM[C]// 2008 Proceedings of the ASME International Design Engi-

neering Technical Conferences and Computers and Information in Engineering Conference. New York, 2008(1):43-51.

- [2] Yuan Changfeng, Wang Wanlei, Li Jianfeng, et al. Approach of knowledge acquisition based on artificial neural network in product design process[C]// 2010 International Conference on Intelligent Computation Technology and Automation, ICICTA. China: Hunan, 2010(1):800-803.
- [3] 罗钧旻. 模糊本体及其演化研究[J]. 微电子学与计算机, 2011, 28(5):105-108.
- [4] Amy J C Trappey, Charles V Trappey. An R and D knowledge management method for patent document summarization[J]. Industrial Management and Data Systems, 2008, 108(2): 245-257.
- [5] David Pinto, Paolo Rosso, Héctor Jiménez-Salazar. A self-enriching methodology for clustering narrow domain short texts[J]. The Computer Journal, 2011, 54(7):1148-1165.
- [6] 赵冬. 知识本体的检索机制研究[J]. 微电子学与计算机, 2011, 28(10):137-145.
- [7] Zhang Dezheng, Zhuang Hongbo. Technique of knowledge acquisition based on domain ontology network model[J]. Computer Engineering, 2007, 61(1):76-92.
- [8] Hays David G. Dependency theory: a formalism and some observations[J]. Language, 1964(40): 511-525.

作者简介:

郭磊 男, (1976—), 硕士, 讲师. 研究方向为本体技术、数据仓库及决策支持系统.



刘运通 男, (1975—), 博士, 讲师. 研究方向为制造业信息化、知识集成技术、本体技术.

