

# 一种跨本体的语义相似度计算方法

黄宏斌 董发花 邓 苏 张维明

(国防科技大学信息系统与管理学院 长沙 410073)

**摘要** 针对在广域分布环境下进行信息共享与服务的需要,本文设计了基于本体的元数据模型,并在 MD3 模型的基础上给出了一种基于该元数据模型的跨本体的语义相似度计算方法。MD3 模型是一种系统的跨本体概念间相似度的计算方法,这种方法无需建立一个集成的共享本体。在 MD3 模型的基础上,充分利用本体对概念的描述信息,重点讨论了跨本体概念间非层次关系相似度的计算,把 MD3 模型扩展到 MD4 模型,使得概念间相似度的计算理论上更全面、更精确。

**关键词** 元数据模型,本体,语义相似度,MD4 模型

## Approach of Determining Semantic Similarity among Concepts between Different Ontologies

HUANG Hong-bin DONG Fa-hua DENG Su ZHANG Wei-ming

(College of Information System and Management, National University of Defense and Technology, Changsha 410073, China)

**Abstract** To meet the demand of sharing information and service in the distribute network, the paper designs the metadata model based on ontology. The MD3 model systematically evaluates semantic similarity across different ontologies dispense with integrating different ontologies into a shared ontology. Based on the MD3 model, the not-hierarchical relations evaluating of concepts from different ontologies is focused on, extended the MD3 model to MD4 model to make the semantic similarity of concepts from different ontologies more comprehensive and precise in theory.

**Keywords** Metadata model, Ontology, Semantic similarity, MD4 model

## 1 引言

随着网络的发展,在一些大型企事业团体和虚拟组织环境中,存在着大量的业务信息系统。组织内的各单位依据业务或专业知识进行分布,并且管理着部分信息资源,不同单位有着不同的工作重心和目标,因此在不同工作领域下对底层的信息资源采用不同的概念实施建模。这些广域网络环境下的信息资源具有动态、分布、多元和无序等特点,各部门或组织之间的信息相对封闭,形成了一个“信息孤岛”,造成了信息资源的浪费。同时,信息资源的异构性为信息之间的共享与互操作带来了困难。对用户而言,访问并有效利用这些信息资源,实现各节点分布信息资源的共享和交换,以及分布信息系统之间的互操作成为一个亟需解决的问题。

本体在人工智能、信息检索、Web 服务发现等领域中扮演着越来越重要的角色,利用本体对信息资源进行建模能够很好反映信息资源的语义。基于本体的元数据通过本体来表示元素模型,将有效地促进领域知识的共享和语义表达,能够为用户和应用提供语义查询和信息汇集能力。但是,现有基于本体的信息共享与服务系统中几乎都假设共享一个集中的全局本体,用来访问信息资源,这样的维护开销比较大。在广域网络环境下,这种假设是不存在的。在网络环境中,每一个结点仅仅维护着本地信息资源的局部领域视图,通过网络中结点的协作来发现、理解和使用领域范围内分布、异构、不断变化的信息资源,异构问题越来越严重,如何解决信息资源的互操作成为一个比较棘手的问题。本体映射和本体集成是解决信息资源语义异构的很好的方法,其中一个关键步骤就是

语义相似度的计算。提高语义相似度计算精度也是提高语义信息检索质量的关键之一。语义相似度一般是指计算本体概念间的相似度,多数方法所考虑的概念是基于一个本体的,跨本体概念间的方法比较少。MD3 模型是一种典型的计算跨本体概念间相似度的方法<sup>[1]</sup>。

本文从现实需求出发,针对在广域分布环境下进行信息共享与服务的需要,针对全局信息视图生成需要设计了基于本体的元数据模型,并在 MD3 模型的基础上给出了一种基于该元数据模型的跨本体的语义相似度计算方法——MD4 模型。

## 2 基于本体的元数据模型

本体(ontology)是领域知识的概念化说明,它将特定领域有关对象、概念及其关系以形式化的说明来严格规定。利用本体建立面向语义的元数据模型,可以将元数据中实体类的含义、类间及对象间的关系更加明确地表达出来,从而支持广域分布环境下的概念建模、信息搜索与交换、信息共享与服务系统建设等研究。

本体在元数据描述和使用的语义化中起着非常重要的作用,主要体现在:1)本体不但能够保证以往基本的元数据描述能力,同时是语义信息描述的基础。语义信息主要由语义类、语义属性、语义关系、语义规则和语义实例构成,这正好对应着本体中的本体概念、概念属性、概念关系、本体规则和公理以及本体实例。2)本体将有效扩展信息检索能力。由于本体描述对各类信息资源赋予了语义,同时本体自身具有一定的推理能力,从而可以利用本体提供面向语义的查询能力,同时

可以利用本体进行查询扩充,从而使检索的结果更加全面。  
3)本体为信息资源提供了一种面向语义的输出能力和相似度计算能力,包括元数据定义的语义输出和信息资源语义实例的语义输出,从而使用户能够更好地理解所需的信息资源。

## 2.1 本体的基本概念

关于 Ontology 的定义有许多,目前获得较多认同的是 R. Studer 的解释<sup>[2]</sup>:“Ontology 是对概念体系的明确的、形式化的、可共享的规范说明”。在最简单的情况下,本体只描述概念的分类层次结构;在复杂的情况下,本体可以在概念分类层次的基础上,加入一组合适的关系、公理、规则来表示概念间的其它关系,约束概念的内涵解释。

**定义 1** 一个完整的本体应由概念、关系、函数、公理和实例等五类基本元素构成。因此,我们把本体表示为如下形式: $O = \{C, R, F, A, I\}$ , 其中:

$C$ :概念。概念是指客观世界中任何事物的抽象描述,在本体中通常按照一定的关系形成一个层次结构。

$R \subseteq 2^{C \times C}$ :概念之间的关系。如“subclass-of”关系、“part-of”关系等。

$F \subseteq R^n$  是一种特殊的关系,其中第  $n$  个元素  $c_n$  相对于前面  $n-1$  个元素是惟一确定的。函数  $F$  可以用如下形式表示: $c_1 \times c_2 \times \dots \times c_{n-1} \rightarrow c_n$ 。

$A$ :概念或者概念之间的关系所满足的公理,是一些永真式。

$I$ :领域内概念实例的集合。

本体用来形式化地描述兴趣领域共享的概念。MD4 模型是 MD3 模型的扩展,在本体的结构以及概念描述两方面做了扩展,主要是增加了非层次关系以及实例两个方面的内容。

## 2.2 基于本体的元数据模型

假设  $T$  为领域语义词典,是领域中术语词汇的集合,主要包括类术语、属性术语、关系术语等, $D_{\text{basic}}$  是预定义的基本数据类型集合, $D_{\text{enum}}$  是预定义的枚举类型。

领域语义词典(Domain semantic dictionary, DSD)用来规范元数据中元素的命名,定义了词汇间的语义关系。DSD 是一个受限的词汇知识库,面向的是领域的常用词汇和专用词汇。其目标是:1)规范不同信息领域中描述信息资源的词汇,从而可以有效地限制任意词汇描述引发的不一致性;2)是中词汇语义相似度计算的基础。DSD 采用《知网》的语义描述结构进行组织,主要对实词进行描述,是一种开放结构,可扩充。

基于信息资源都由信息单元组成这一理解,采用信息单元实体类和实体类之间的关系以及实体对象、相关约束和规范描述信息资源。具体定义如下:

**定义 2** 元数据模型是一个八元组  $MD = \langle E, A, L, H^C, R, I, F, P \rangle$ , 其中:

1)  $E$  是信息单元实体类集合,  $\forall c \in E, c = (\text{Name}, A^c)$ ,  $\text{Name}(c)$  表示  $c$  的命名词汇,  $\text{Name}(c) \in T$ , 其中  $A^c = \{x \mid (x \in A) \cap (\text{Att}(x) = c)\}$ 。  $\text{Att}$  是函数集中的属性映射函数。

2)  $A$  是信息单元实体类的属性集合,属性又分为基础属性和复合属性。 $\forall a \in A, a = (\text{Name}, dt)$ ,  $\text{Name}(a)$  表示  $a$  的命名词汇,  $\text{Name}(a) \in T, dt \in D_{\text{basic}} \cup D_{\text{enum}}$ 。

3)  $L$  取值域集,  $L = D_{\text{basic}} \cup D_{\text{enum}}$ ;

4)  $H^C$  表示类间的二元层次关系,层次关系是有向传递的偏序关系,包括了类间的继承 is-a 关系和聚合(组合)part-

of 关系。 $H^C \subseteq E \times E, H^C = H^{\text{Inh}} \cup H^{\text{Agr}}$ ,  $H^{\text{Inh}}(c_1, c_2)$  表示  $c_2$  是  $c_1$  的子类,  $H^{\text{Agr}}(c_1, c_2)$  表示  $c_2$  是  $c_1$  的组成部分,如:  $H^{\text{Inh}}$ (学员,本科生),  $H^{\text{Agr}}$ (学院,系)。

5)  $R$  是实体类之间的二元关系集。任何所连接的实体类超过两个的关系,都能够转换为一组二元多对关系集合。因此,该模型中的关系设计为二元关系。 $\forall r \in R$  可表示为  $N^r(c_1, c_2), c_1, c_2 \in E, N^r \in T$  为关系的名称。如  $R(\text{授课}) = (\text{教员}, \text{学员})$ 。

6)  $I$  实例集,是信息单元实体类对象集合。

7)  $F$  表示函数集,主要包括如下函数:

$\text{Att}: A \rightarrow E \cup R$ , 属性函数,将属性分配给某一实体类或关系;

$\text{val}: V \rightarrow (E, A)$ , 属性取值函数;

$\text{Inst}: E \rightarrow 2^I$ , 信息单元实例化函数,可以写为:  $\text{Inst}(E) = I$  或  $E(I)^{[3]}$ ;

$\text{Instr}: R \rightarrow 2^{I \times I}$ , 关系实例化函数,可以写为:  $\text{Instr}(R) = \{I_1, I_2\}$  或  $R(I_1, I_2)^{[3]}$ ;

$\text{dom}: R \rightarrow E$ , 由  $\text{dom}(R) = \Pi_1(\text{rel}(R))$  给出其定义域;

$\text{range}: R \rightarrow E$ , 由  $\text{range}(R) = \Pi_2(\text{rel}(R))$  给出其值域。

8)  $P$  约束规则集,所有信息单元实体类和关系要满足的约束的集合。包括取值约束和基数约束。

**定义 3** 给定  $MD = \langle E, A, H^C, R, L, I, F \rangle$ ; 语义解释为一个三元组,记为  $I = \langle \Delta^I, \Delta^L, \cdot^I \rangle$ , 其中  $\Delta^I$  是一个非空集合,包含领域中的所有个体; $\Delta^L$  是一个非空集合,包含领域中的所有数据值; $\cdot^I$  是解释函数,它将  $E$  中的每个实体类  $C$  都映射为  $\Delta^I$  的一个子集  $C^I (C^I \subseteq \Delta^I)$ , 每个具体域  $L$  解释成一个集合  $L^I (L^I \subseteq \Delta^L)$ , 每个关系  $R$  解释成一个二元关系  $R^I (R^I \subseteq \Delta^I \times \Delta^I)$ , 每个属性  $A$  解释成一个二元关系  $A^I (A^I \subseteq \Delta^I \times \Delta^L)$ , 每个个体  $c$  解释成一个元素  $c^I (c^I \in \Delta^I)$ , 每个值  $l$  解释成一个元素  $l^I (l^I \in \Delta^L)$ 。

## 3 常用语义相似度计算方法

对于语义相似度计算不同领域有不同的方法,主要分为基于特征的模型、基于语义距离的模型、基于信息内容的模型以及混合模型。

### 3.1 基于特征匹配的模型

Tversky 模型是典型的特征匹配模型,提出了基于集合理论,用特征匹配过程来计算相似度的方法<sup>[4]</sup>。指出相似度不仅由两个概念的相同属性决定,而且由它们的不同属性决定。这种模型没有考虑属性与概念间的密切程度。由于该模型符合信息领域对相似度的认知,在信息领域有众多学者在 Tversky 模型的基础上建立语义相似度模型,如文献<sup>[5-7]</sup>等。

### 3.2 基于语义距离的模型

基于语义距离的模型根据概念在层次结构中的位置来计算语义相似度。层次结构中结点表示概念,边表示语义关系。在概念层次树中,任何两个结点之间有且只有一条最短路径,把这条路径的长度作为两个概念的语义距离的度量。该模型适合很多专业领域,但它过分依赖事先定义好的语义网结构,只考虑了概念间的聚合(IS-A)关系,计算出来的有相同父类的子类之间的相似度值很粗糙。

### 3.3 基于信息内容的方法

基于信息内容的方法<sup>[6]</sup>不依赖于层次结构的词典,只要有概率模型存在,这种方法就可以应用。两个概念的相似性决定于它们共享信息的程度,共享信息含量越大,两个概念越

相似。两个概念的共享信息含量用两个概念在概念树上的所有共同“超类”所具有的最大信息含量来表示。

### 3.4 混合模型

Matching-Distance(MD)模型是一种典型的混合模型<sup>[8]</sup>,该模型以 Tversky 的比率模型<sup>[4]</sup>为基础,其中特征权重是根据实体之间的语义关系来确定的。MD 模型综合了基于特征匹配和基于语义距离两种方法,该方法在语言学上有很深的根基,定义了非对称的语义相似度计算函数。

### 4 MD3(Triple Matching-Distance Model)模型

概念间相似度的计算方法根据要比较的概念是否来自同一个本体分为单本体和跨本体概念相似度计算方法。基于语义距离和基于信息内容的方法分别利用了本体的结构信息(概念术语的位置信息)和信息内容,不同本体的结构和信息内容不能直接进行比较,这两种方法适合单本体概念相似度的计算。跨本体的相似度方法通常使用混合的或基于特征的方法。

跨本体相似度计算大都采用综合的相似度计算方法,分别计算概念的名称、属性以及概念间语义关系间的相似度,然后通过加权求和得到概念之间的相似度。其中 MD3 模型是一种典型的计算跨本体概念间语义相似度的方法,是 MD 模型在跨本体应用中的扩展。

MD3 模型是一种跨本体概念间相似度计算框架。计算实体类  $a$  和  $b$  之间的相似度通过计算同义词集、特征属性和语义邻居之间的加权和,公式如下:

$$Sim(a, b) = wS_{synsets}(a, b) + uS_{features}(a, b) + vS_{neighborhoods}(a, b)$$

其中  $w, u, v$  表示了各组成部分的重要性。特征属性细化为组成部分、功能以及其他属性。概念  $a$  和  $b$  的语义邻居及其特征属性(即概念的部分、功能及其他属性)也通过同义词集合描述,每一个相似度的计算都通过 Tversky<sup>[4]</sup>公式:

$$S(a, b) = \frac{|A \cap B|}{|A \cap B| + \alpha(a, b)|A - B| + (1 - \alpha(a, b))|B - A|}$$

其中  $A, B$  分别表示概念  $a$  和  $b$  的描述集合,  $A - B$  表示属于  $A$  但不属于  $B$  的术语集( $B - A$  相反)。参数  $\alpha(a, b)$  由概念  $a$  和  $b$  在各自层次结构中的深度确定,公式如下:

$$\alpha(a, b) = \begin{cases} \frac{\text{depth}(a)}{\text{depth}(a) + \text{depth}(b)}, & \text{depth}(a) \leq \text{depth}(b) \\ 1 - \frac{\text{depth}(a)}{\text{depth}(a) + \text{depth}(b)}, & \text{depth}(a) > \text{depth}(b) \end{cases}$$

### 5 Fourfold Matching-Distance Model(MD4)模型

传统的跨本体概念间相似度的计算通过手工或半自动方式把不同本体集成为一个共享本体,然后把本地本体中的概念映射到这个共享本体,再利用单本体概念间相似度的计算方法来完成跨本体概念间的比较。在某些情况下,这种集成共享本体的方法是不现实的,或者形成这个共享本体的代价很大,必须考虑直接针对两个不同本体来比较其概念。MD3 模型<sup>[1]</sup>用一个通用的虚拟的根把两个本体联系起来,分别计算概念名称、特征属性以及语义邻居之间的相似度。其不足在于其语义邻居只考虑了层次语义关系,没有考虑语义关系中非层次关系的影响,同时对对象实例对于概念的影响也没有考虑。本文在 MD3 模型的基础上,参考了其概念名称相似度、特征属性,重点讨论了跨本体概念间非层次关系的相似度的比较和实例对概念相似度的影响,把 MD3 模型扩展到

MD4 模型。

概念由三部分组成:概念名、概念的属性以及概念间的语义关系。概念由同义词集合描述。概念的属性细化为功能、部分以及其他属性。概念的语义关系包含两部分:层次语义关系和非层次语义关系。其中最常见层次语义关系是上下位关系和部分整体关系。这类关系是有向传递的非对称关系。而定义的非层次关系不具有传递性,因此计算概念间相似度需对这两类关系分别讨论。

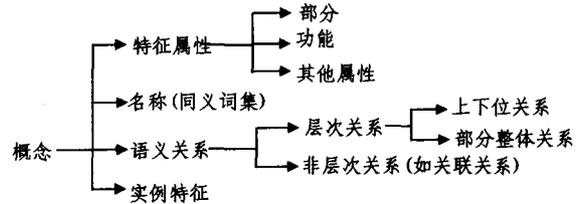


图1 基于本体元数据模型的概念描述组成

#### 5.1 概念名称相似度

本文元数据模型中所有实体名称均是在领域语义词典中定义的,因此实体名称相似度计算可以采用基于知网的概念距离。知网<sup>[9]</sup>中概念的语义用义原来描述,义原是描述概念语义的最小单位,一共采用了 1500 多个义原。这些义原的作用并不是等同的,而是存在复杂的关系。为了简单起见,这里仅考虑义原的上下位关系。根据这种关系,所有义原构成了一个树状的层次体系。假如两个义原在该层次体系中的路径为  $d$ ,可得到两个义原之间的语义相似度,如下式所示:

$$Sim(p_1, p_2) = \frac{a}{d+a}$$

上式中  $a$  是一个可调节的参数。由于概念的语义是用多个义原来表示的,因此我们只要计算每个义原的相似度来考虑其重要性,就可以得到概念之间的名称相似度。计算方法如下:

$$S_{name}(c_1, c_2) = \sum_{i=1}^m w_i \text{Max}_{1 \leq j \leq n} Sim(p_i, p_j)$$

其中,  $m, n$  为概念  $c_1, c_2$  的义原数,  $w_i$  为第  $i$  个义原所占的权重。

#### 5.2 概念属性的相似度

文献<sup>[1]</sup>中,特征属性被细化为三类:组成部分、功能以及其他属性。这样做的优点是可以依据语言环境的不同,根据属性对概念相似度影响的大小不同分配不同的权值。把三部分属性相似度的加权和作为概念特征属性的相似度,公式如下:

$$S_u(a^p, b^q) = w_1 S_1(a^p, b^q) + w_2 S_2(a^p, b^q) + w_3 S_3(a^p, b^q)$$

其中  $S_1, S_2, S_3$  分别是组成部分、功能和其他属性的相似度,其相应的权值为  $w_i$ ,并且满足  $w_i \geq 0, \sum_{i=1}^3 w_i = 1$ 。

#### 5.3 语义关系的相似度

语义关系包括层次语义关系和非层次语义关系,层次关系是有向传递的,非层次语义关系不具有传递性(如关联关系)。

##### (1) 层次语义关系的计算

计算层次语义关系借鉴文献<sup>[1]</sup>中语义邻居的定义,以概念为中心向周围辐射。设定一个语义半径,半径取值的大小反映与概念之间的亲疏关系。划定语义邻居的范围集合,进行匹配,取集合中的最大值作为语义邻居之间的相似度,公式如下:

$$N(a^o, r) = \{c_i^o \mid \forall i, d(a^o, c_i^o) \leq r\}$$

(2)非层次语义关系的计算

上位词:定义概念的上位词为概念所有父类的集合,公式如下:

$$UC(C_i, H) = \{C_j \in C \mid H(C_i, C_j)\}$$

基于概念上位词的定义,定义概念的匹配公式如下:

$$CM(C_1, O_1; C_2, O_2) = \frac{|UC(C_1, H_1) \cap UC(C_2, H_2)|}{|UC(C_1, H_1) \cup UC(C_2, H_2)|}$$

与概念相关的非层次关系:如果关系的定义域或值域是概念  $c$ ,则称这些关系为与概念  $c$  相关的非层次关系,公式如下:

$$R_c(P) = \{\text{dom}(R_x) = c \cup \text{range}(R_x) = c \mid R_x \in P, c \in C\}$$

还可进一步把非层次关系细化为概念的 IN 关系和 OUT 关系(可以认为非层次关系的方向是从定义域到值域,凭此来定义 IN 和 OUT 关系),IN 关系指概念  $c$  是非层次关系的值域,公式如下:

$$R_{c-I} = \{\text{range}(R_x) = c \mid R_x \in P, c \in C\}$$

而 OUT 关系指概念  $c$  是非层次关系的定义域,公式如下:

$$R_{c-O} = \{\text{dom}(R_x) = c \mid R_x \in P, c \in C\}$$

比较概念的非层次关系,首先应找出两个本体中,与这两个概念相关的同类非层次关系(无需考虑不同类的非层次关系),进而比较这些同类非层次关系的另外一项之间的相似度(如果要比较的概念是非层次关系的定义域,分别找出这个关系的值域,通过概念匹配公式对其进行比较,反之亦然)。

下面以 IN 关系为例描述比较的过程(OUT 关系类似):

$$R_I = R_{a-I}^p \cap R_{b-I}^q$$

其中  $R_{a-I}^p$  表示本体  $p$  中与概念  $a$  相关的 IN 关系,而  $R_{b-I}^q$  表示本体  $q$  中与概念  $b$  相关的 IN 关系,所以其交集  $R_I$  表示本体  $p, q$  中与概念  $a, b$  相关的公共 IN 关系集合。如果概念  $a, b$  没有公共的 IN 关系,则  $R_I$  为空,无需下面的计算。对于公共 IN 关系集合,公式如下:

$$S_{mm-h-1}(a, b) = \frac{1}{|R_I|} \sum_{i=1}^{|R_I|} CM(\text{dom}(R_{I_i}), p; \text{dom}(R_{I_i}), q)$$

对 IN 关系和 OUT 关系进行加权综合,得到非层次关系相似度的公式如下:

$$S_{mm-h}(a, b) = iS_{mm-h-1}(a, b) + oS_{mm-h-O}(a, b)$$

其中  $i, o$  为权值,反映了不同类型关系对概念相似度的影响。对层次关系以及非层次关系计算结果进行综合,得到概念语义环境的相似度,公式如下:

$$S_{neighborhoods}(a, b) = tS_h(a, b) + uS_{mm-h}(a, b)$$

5.4 实例特征相似度

概念  $a, b$  的实例特征相似度是通过资源元数据描述中具体实例和实例特征来评估,概念  $a, b$  实例特征相似度计算函数为

$$S_{inst}(a, b) = \frac{P(a \cap b)}{P(a \cup b)} = \frac{P(a, b)}{P(a, b) + P(a, \bar{b}) + P(\bar{a}, b)}$$

基于实例特征计算概念相似度牵扯到三个概率:  $P(a, b), P(a, \bar{b}), P(\bar{a}, b)$ 。其中,  $P(a, b)$  是从一个本体的实例空间

中随机选取一个实例属于概念  $a$ , 并且同时属于概念  $b$  的实例在实例空间中所占的比重。

5.5 综合公式

基于上面的分析,最后综合四部分的相似度值<sup>[10]</sup>,得到跨本体概念间相似度的综合公式如公式如下:

$$Sim(a, b) = wS_{name}(a, b) + uS_{features}(a, b) + vS_{neighborhoods}(a, b) + rS_{inst}(a, b)$$

6 性能分析

概念相似度计算方法的评估通常是把计算结果与人主观判断结果进行比较,两者越接近,说明概念相似度方法就越好。由于知识背景和认知经验不同,判断结果主观性很强。即使取一群人判断结果的平均值,也只能相对客观地反映这一群人的认知特性,不能客观反映概念之间的相似性。因此,如何客观地评估概念相似度的计算方法,还有待于进行深入的研究。

本文提出的 MD4 模型继承了 MD3 模型的优点,对 MD3 模型进行了完善。在描述丰富的本体中,通过选择合适的权值,MD4 模型理论上可以确保概念相似度的计算更全面、准确。当然,在概念相似度计算过程中存在大量权值的设定,对性能存在一定的影响。如何方便快捷地选择合适的权值是未来工作的重点。

参考文献

- [1] Rodriguez M A, Egenhofer M J. Determining Semantic Similarity Among Entity Classes from Different Ontologies. IEEE Trans. on Knowledge and Data Engineering, 2003, 15(2): 442-456
- [2] Studer R, et al. Knowledge engineering, principles and methods. Data and Knowledge Engineering, 1998, 25: 161-197
- [3] Maedche A. Clustering Ontology-based Metadata in the Semantic Web
- [4] Tversky A. Features of similarity. Psychological Review, 1977, 84(4): 327-352
- [5] Lee J, Kim M, Lee Y. Information retrieval based on conceptual distance in IS-A hierarchies. Journal of Documentation, 1993, 49(2): 188-207
- [6] Resnik O. Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity and natural language. Journal of Artificial Intelligence Research, 1999, 11: 95-130
- [7] Rada R, Mili H, Bicknell E, et al. Development and application of a metric on semantic nets. IEEE Transactions on System, Man, and Cybernetics, 1989, 19(1): 17-30
- [8] Rodriguez M A. Assessing Semantic Similarity among Spatial Entity Classes. PhD thesis. University of Maine, 2000
- [9] 刘群, 李素建. 基于《知网》的词汇语义相似度计算[J]. 中文计算语言学, 2002, 7(2): 59-76
- [10] 程勇, 黄河, 等. 一个基于相似度计算的动态多维概念映射算法[J]. 小型微型计算机系统, 2006(6)