

# 基于本体结构的语义相似度计算

杨方颖, 蒋正翔, 张姗姗  
(北京理工大学 计算机学院, 北京 100081)

**摘 要:**语义相似度是语义网络和信息检索领域的重要内容。本体结构为语义相似度计算提供了新的思路,但现有的方法都存在着不同程度的缺陷。为了提高已有方法的有效性,在分析语义相似度经典方法的基础上,充分利用本体的结构信息,综合考虑概念在本体图中的位置、语义距离,共享属性量和共享信息等因素,提出了一个基于本体结构的语义相似度算法。实验部分以维基百科中公开发布的氨基酸本体为例,通过与经典方法计算结果的对比,证明了算法的有效性。

**关键词:**本体;语义相似度;语义距离;信息内容

**中图分类号:**TP18

**文献标识码:**A

**文章编号:**1673-629X(2013)07-0052-05

doi:10.3969/j.issn.1673-629X.2013.07.013

## Semantic Similarity Measurement Based on Ontology

YANG Fang-ying, JIANG Zheng-xiang, ZHANG Shan-shan  
(College of Computer, Beijing Institute of Technology, Beijing 100081, China)

**Abstract:** Semantic similarity is one of the most important parts in domains of semantic networks and information retrieval. The structure of ontology provides a new perspective for semantic similarity measurement. However, there are varying degrees of defaults in the existing methods. In order to improve the effectiveness of the existing methods, a novel approach for semantic similarity measurement based on ontology construction is proposed after a deep research on various classical approaches. It takes into account many factors, including semantic distance, level, semantic coincidence degree and information content. In order to verify the effectiveness of the algorithm, some comparative experiments are conducted using the Wikipedia disclosed amino acid ontology as an example.

**Key words:** ontology; semantic similarity; semantic distance; information content

### 0 引言

目前语义相似度计算已经广泛地用于自动问答、文本分类、信息检索、数据挖掘和自然语言处理等领域,是当前信息领域技术研究的热点。目前基于本体的相似度算法主要分为两类<sup>[1]</sup>:基于边的算法和基于顶点的算法。基于边的算法通过统计本体层次图中两概念之间的边数来衡量二者的语义距离。该算法认为,语义距离越大,相似度越低。基于顶点的算法根据节点的信息量来计算节点语义相似度,该类算法认为,本体中概念出现的概率越大,其携带的信息量就越少。上述两种算法都存在着不同程度的缺陷。基于边的算法缺陷源于该算法的两个假设:节点和边在本体图中均匀分布;同一本体中,如果两对概念的最短距离相同,那么其语义距离也相同。而这两个假设在大多数的情况下是不成立的。基于顶点的方法在一定程度上

弥补了基于边方法的不足,但是它只考虑了两概念共同祖先节点包含的信息量,而忽略了概念之间的距离,影响了语义相似度的准确性。除此之外,两个算法都没有考虑到节点在本体图中的层次和属性等特征。为了克服两者的不足,文中提出了一种新的算法,本算法吸取了上述两种算法的优点,并且加入了层次特征和属性特征的度量,最后通过实验的分析对比表明,该算法在计算语义相似度方面有更高的准确性。

### 1 相关工作

词语的语义相似度<sup>[2]</sup>主要用于衡量文本中词语的可替换程度。在不改变原文的句法和语义结构的情况下,词语的可替换程度越高,语义相似度就越高。

基于边的方法是最早发展起来的。Resnik<sup>[3]</sup>提出了一种将最短路径长度转换为相似度的方法:

收稿日期:2012-10-08

修回日期:2013-01-12

网络出版时间:2013-04-08

基金项目:国家自然科学基金资助项目(61203452)

作者简介:杨方颖(1989-),女,山东菏泽人,硕士研究生,研究方向为人工智能、自然语言处理。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20130408.1607.038.html>

$$\text{Similarity}(a, b) = (2 \times D) - \text{Len}(a, b) \quad (1)$$

其中  $a, b$  分别为本体图中的两个概念,  $\text{Len}(a, b)$  为  $a, b$  的最短路径长度,  $D$  为本体图的最大深度。该算法的创新之处在于,它不仅考虑了语义距离对相似度的影响,还加入了本体图深度因素。后来, Leacock<sup>[4]</sup>使用  $\log$  函数改进了该算法模型:

$$\text{Similarity}(a, b) = -\log\left(\frac{\text{Len}(a, b)}{2 \times D}\right) \quad (2)$$

以上两种算法模型都没有对计算结果做归一化处理。刘群<sup>[2]</sup>等使用模型

$$\text{Similarity}(a, b) = \frac{\alpha}{\alpha + \text{Len}(a, b)} \quad (3)$$

来计算网中义原之间的相似度,该模型的优点是简单易用,结果归一化,缺点是只考虑了节点的语义距离。

如引言所述,为了克服基于边的算法带来的缺陷, Resnik<sup>[3]</sup>提出了基于顶点的语义相似度算法。他将概念的相似度定义为其最近共同祖先(LCA, the lowest common ancestors)的信息量(IC, information content)的最大值。

$$\text{Similarity}(a, b) = \max_{c \in \text{LCA}(a, b)} IC(c) \quad (4)$$

其中信息量  $IC(c)$  为概念  $c$  出现概率的负  $\log$  函数值:  $IC(c) = -\log P(c)$ 。

其中  $P(c) = \frac{n(c)}{N}$ ,  $n(c)$  为概念  $c$  所包含的子概念数,  $N$  为本体中概念的总数。该算法忽略了两个节点本身的信息。只关注其最近共同祖先节点,这将导致同一最近祖先的所有概念无论其本身在本体图中的位置如何,都将有相同的相似度,而这显然是不合理的。Lin<sup>[5]</sup>在信息量的基础之上,将相似度定义为描述概念的共性所需的信息量和完全描述两个概念所需信息量的比值:

$$\text{Similarity}(a, b) = \frac{2 \times IC(\text{LCA}(a, b))}{IC(a) + IC(b)} \quad (5)$$

该算法加入了两个概念节点本身的信息,在一定程度上改善了公式(4)的不足,但是它在两节点的祖先节点中,仍然只关心最近祖先节点的信息。Couto<sup>[6]</sup>在2005年提出了分离共同祖先的概念,并在其算法模型 GraSM 中首次使用分离共同祖先的信息量来代替最近共同祖先的信息量。一个节点有分离祖先代表该节点有两种以上的语义解释。通过使用共同分离祖先来代替最近共同祖先,该模型的最大优势在于它加入了概念节点的其他可能的语义解释。2011年 Couto<sup>[7]</sup>修改了 GraSM 算法,提出了新的模型 DiShIn, DiShIn 相比 GraSM 在算法效率上以及多重解释(如果一个概念有两个以上的直接父类,则该概念有多重解释)的管理上有较大提高。由于 DiShIn 算法的优越性,文中

在计算信息量度量值时选用了该算法。DiShIn 算法的内容将在下文中详细说明。

除了以上的两种主要方法外,还有将这两种方法结合起来的混合方法等。其他研究包括: Tversky<sup>[8]</sup>从属性的角度出发比较两个概念的相似度。他认为,事物是由属性特征来反映的,事物之间的关联程度与其共同属性数有很大关系。林智超<sup>[9]</sup>等通过对概念属性建立网络流最小费用最大流模型来计算概念的相似度。张晓李<sup>[10]</sup>等在概念属性的基础上引入了相似图的结构。赵小谦<sup>[11]</sup>等还考虑了概念在本体图中的密度对相似度的影响。段寿建等<sup>[12]</sup>综合考虑了概念所在层次、语义距离以及语义重合度等特征提出了自己的相似度的算法。Rodriguez 等<sup>[13]</sup>还考虑了概念的同义词集合,区别特征和语义相邻点等特征。

文中结合了基于边的方法和基于顶点的方法,同时吸取了 Tversky<sup>[8]</sup>关于属性的观点,综合考虑了语义距离、属性、共同父节点所在层次、信息量等影响语义相似度的特征,提出了自己的相似度计算方法,下面将详细说明。

## 2 算法介绍

文中将距离因子  $D$ , 层次因子  $L$ , 属性因子  $P$  和信息量因子  $IF$  通过线性加权的方式组合起来,得到了一个新的语义相似度算法,本算法形式化表示如下:

$$\text{Similarity}(c_1, c_2) = \begin{cases} 1 & (c_1 = c_2) \\ \alpha D + \beta L + \gamma P + \theta IF & (c_1 \neq c_2) \end{cases} \quad (6)$$

其中  $\alpha + \beta + \gamma + \theta = 1$ ,  $D \in [0, 1]$ ,  $L \in [0, 1]$ ,  $P \in [0, 1]$ ,  $IF \in [0, 1]$ 。

通过算法的定义可知:

1)  $\text{Similarity}(c_1, c_2) \in [0, 1]$ ;

2) 当  $c_1$  和  $c_2$  是等价概念时,其相似度取得最大值为1;

3) 各个特征因子取值越大,相似度越大。

算法四个特征因子的定义和选取依据如下:

### 2.1 距离因子 $D$

语义距离是衡量相似度的一个重要指标。语义距离越大,概念的语义差异越大,语义相似度越小。这里使用距离因子  $D$  来表示语义距离对相似度的影响。

语义距离:两个概念  $c_1, c_2$  的语义距离为:在本体层次图两概念的任意最近共同祖先节点  $p$  分别到  $c_1$  和  $c_2$  的最短路径长度之和。

$$\text{Distance}(c_1, c_2) = \text{edge\_count}(p, c_1) + \text{edge\_count}(p, c_2), \forall p \in \text{LCA}(c_1, c_2) \quad (7)$$

其中  $\text{edge\_count}(p, c_1)$  表示本体层次图中从  $p$  到

$c_1$  的最短路径中边的个数(即最短路径长度)。如图 1 所示,概念节点  $j, k$  的共同祖先有:  $h, g, f, a, n, c, b$ 。由于节点  $h$  和节点  $n$  到两节点的最短距离之和均为最小,因此  $LCA(j, k) = \{h, n\}$ 。

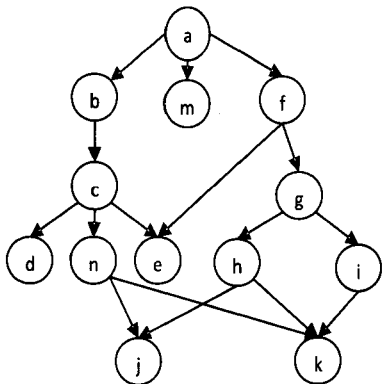


图 1 本体片段示意图

$$\begin{aligned} \text{Distance}(j, k) &= \text{edge\_count}(h, j) + \text{edge\_count}(h, k) \\ &= \text{edge\_count}(n, j) + \text{edge\_count}(n, k) = 2 \end{aligned}$$

距离因子  $D$ : 用于度量语义距离对语义相似度的影响。考虑到计算复杂度和归一化需要,使用刘群<sup>[2]</sup>在计算知网中义原相似度时的计算模型作为本算法的距离因子:

$$D(c_1, c_2) = \frac{\lambda}{\lambda + \text{Distance}(c_1, c_2)} \quad (8)$$

## 2.2 层次因子 $L$

概念在本体图中所处的层次反映了概念的具体化程度。层次越高,概念细化程度越高,其子概念的差异程度也就越小。

概念节点  $c$  的层次  $\text{Level}(c)$  定义为:在本体图中,从唯一的顶级概念节点  $\text{thing}$ (在图 1 中为节点  $a$ ) 到节点  $c$  的最短路径长度加 1。

$$\text{Level}(c) = \text{edge\_count}(\text{thing}, c) + 1 \quad (9)$$

如图 1 所示,  $\text{Level}(a) = 1, \text{Level}(e) = 3$ 。

概念的共同祖先代表概念的共同解释,是概念共性的体现。这里使用概念的共同祖先计算两概念的层次因子( $\text{Level}_{\text{parent}}(c_1, c_2)$ )。本体图是个有向无环图,两个概念可能存在不止一个最近共同祖先,取这些最近共同祖先在本体图中所处层次的最小值。即:

$$\begin{aligned} \text{Level}_{\text{parent}}(c_1, c_2) &= \min \{ \text{Level}(a) \}, a \in \\ &\text{LCA}(c_1, c_2) \end{aligned} \quad (10)$$

如图 2 所示,  $\text{LCA}(f, g) = \{e, c\}$ , 由于  $\text{Level}(e) = 3, \text{Level}(c) = 2$ , 因此  $\text{Level}_{\text{parent}}(f, g) = \min \{ \text{Level}(e), \text{Level}(c) \} = 2$ 。

同样地,本体图的深度反映了整个本体的具体化程度。定义本体图的深度( $\text{Depth}$ )为本体图中从顶层节点  $\text{thing}$ (图 2 中为节点  $a$ ) 到其他任意节点的简单路径长度的最大值加 1,即:

$$\text{Depth} = \max \{ \text{MaxSimplePathLen}(\text{thing}, c) \} + 1 \quad (11)$$

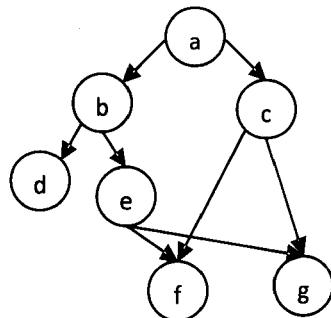


图 2 本体片段示意图 2

其中  $c$  为本体层次图中除了唯一顶层节点  $\text{thing}$ (在图 2 中是节点  $a$ ) 以外的其他任意节点,  $\text{MaxSimplePathLen}(\text{thing}, c)$  表示从节点  $\text{thing}$  到节点  $c$  的所有简单路径中路径长度的最大值。如图 2 所示,  $\text{MaxSimplePathLen}(a, g) = 3$ (即路径  $a - b - e - g$  的长度),该本体图深度为 4。

为了方便对层次因子进行归一化处理,采用商的形式来定义层次因子:

$$L(c_1, c_2) = \frac{\text{Level}_{\text{parent}}(c_1, c_2)}{\text{Depth}} \quad (12)$$

## 2.3 属性因子 $P$

概念的性质是通过其属性来刻画的,共同属性占的比例越大,相似的性质就越多,其相似度也就越大。因此使用属性因子  $P$  来描述共同属性对相似度的影响。

概念  $c_1, c_2$  的属性  $P(c_1, c_2)$  为  $c_1, c_2$  的共同属性在其所有属性中所占的比例。

$$P(c_1, c_2) = \frac{|\text{property}(c_1) \cap \text{property}(c_2)|}{|\text{property}(c_1) \cup \text{property}(c_2)|} \quad (13)$$

其中  $\text{property}(c_1)$  表示  $c_1$  的属性集合,  $|X|$  表示集合  $X$  的元素个数,  $\cap$  表示集合的交集,  $\cup$  表示集合的并集。

## 2.4 信息量因子 $IF$

信息量从概率的角度衡量了概念语义信息的大小。它克服了语义距离忽视概念描述能力的缺陷。在这里,沿用公式(5)的定义,将概念  $c_1, c_2$  的信息量因子  $IF(c_1, c_2)$  定义如下:

$$IF(c_1, c_2) = \frac{2 \times IC_{\text{DiShIn}}(\text{LCA}(a, b))}{IC(a) + IC(b)} \quad (14)$$

它与公式(5)不同的是,分子部分采用在计算祖先节点信息量时表现更为出色的  $\text{DiShIn}^{[7]}$  模型。使用  $\text{DiShIn}^{[7]}$  模型计算  $IC_{\text{DiShIn}}(\text{LCA}(a, b))$  的步骤如下:

- 1) 找出  $IC_{\text{DiShIn}}(\text{LCA}(a, b))$  的所有共同祖先;
- 2) 对每个共同祖先  $\text{parent}$ , 分别计算其到概念  $c_1, c_2$  的独立路径数目,并将独立路径条数分别记为  $m_1,$

$m_2$ , 将  $m_1$  和  $m_2$  差的绝对值  $(|m_1 - m_2|)$  作为祖先 parent 的路径特征 feature。这样就得到了若干个由祖先  $parent_i$  和特征  $feature_i$  组成的二元组  $\langle parent_i, feature_i \rangle$ ;

3) 比较这些二元组的  $feature_i$  的值, 对于  $feature_i$  相同的若干个二元组, 只保留  $IC(parent)$  最大的那个; 将处理完成后剩余的二元组中的 parent 节点组成集合  $c(parent)$ ;

4)  $IC_{DisShn}(LCA(a,b)) = \text{avg}\{IC(parent_i)\}, parent_i \in C(parent)$ 。

综上所述, 综合考虑了概念在本体图中的距离因子、层次因子、属性因子和信息量因子对语义相似度的影响, 提出了全新的基于本体结构的语义相似度算法。下面将通过对比实验说明本算法的有效性。

3 实验部分

3.1 数据的获取

目前领域本体有很多, 以维基百科中公开发布的氨基酸本体为例, 比较算法 (3) (刘群<sup>[2]</sup>), 算法 (1) (Resnik<sup>[3]</sup>), 算法 (5) (Lin<sup>[5]</sup>) 和本算法的性能。部分氨基酸本体结构如图 3 所示。

同时为了方便比较各个算法的性能, 选取了一些有代表性的概念对。概念对和选取依据如表 1 所示。

3.2 评测标准

语义相似度是一个较为主观的概念, 由于语言的复杂性, 很难将两个词语的相似或者差异程度使用准

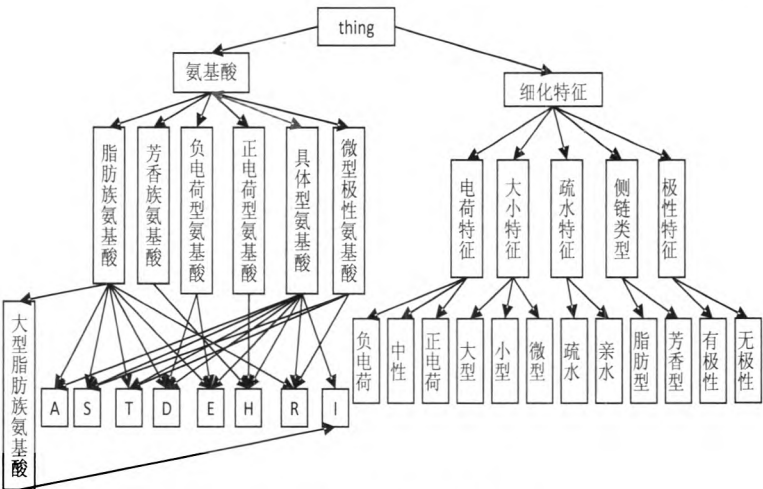


图 3 部分氨基酸本体图

确的数值表达出来。这里, 使用领域专家在  $[0,1]$  的范围内给出的各个概念对的相似度经验值作为语义相似度的评测标准, 认为越接近领域专家的经验值, 算法的准确度越高。

3.3 实验结果

在维基百科公开发布的氨基酸本体的基础上, 分别计算了各个概念对在公式 (3) (刘群<sup>[2]</sup>), 公式 (1) (Resnik<sup>[3]</sup>), 公式 (5) (Lin<sup>[5]</sup>) 和文中算法 (公式 (6)) 下的相似度结果, 并在表格的最右边给出领域专家的经验值作为对比。为了统一相似度的取值范围, 方便比较, 对于公式 (1) 进行归一化处理, 归一化后计算公式为:  $Similarity(c_1, c_2) = \frac{2 \times D - Len(c_1, c_2)}{2 \times D}$ 。本算

法的参数值设置为:  $\alpha = \frac{1}{7}, \beta = \frac{1}{7}, \gamma = \frac{5}{14}, \theta = \frac{5}{14}\lambda = 0.1$ , 实验结果如表 1 所示。

表 1 选取依据及实验结果

序号	概念 $c_1$	概念 $c_2$	期望值说明	算法 (3)	算法 (1)	算法 (5)	文中算法	期望值
1	氨基酸	细化特征	顶层概念, 相似度较小	0.0476	0.8	0	0.0354	0.03
2	氨基酸	侧链类型	相似度比 1 更小	0.0323	0.7	0	0.0322	0.02
3	氨基酸	有极性	相似度比 2 更小	0.0244	0.6	0	0.0321	0.01
4	thing	氨基酸	父子概念	0.0909	0.9	0	0.0416	0.08
5	氨基酸	脂肪族氨基酸	共同子类较多	0.0909	0.9	0.7870	0.4226	0.48
6	氨基酸	微型极性氨基酸	相同子类较少	0.0909	0.9	0.3353	0.3327	0.27
7	氨基酸	具体氨基酸	等价概念	1	1	1	0.9143	1.00
8	氨基酸 S	氨基酸 T	属性相同	0.0476	0.8	0.7147	0.7049	0.83
9	氨基酸 D	氨基酸 E	属性相同	0.0476	0.8	0.7147	0.7049	0.83
10	氨基酸 A	氨基酸 H	属性全部不同	0.0476	0.8	0.1440	0.1154	0.20
11	氨基酸 A	氨基酸 R	1 个属性不同	0.0476	0.8	0.2219	0.2293	0.25
12	负电荷型氨基酸	脂肪族氨基酸	共同子类较多	0.0476	0.8	0.3074	0.1737	0.31
13	负电荷型氨基酸	正电荷型	无共同子类	0.0476	0.8	0.2126	0.1399	0.23
14	thing	氨基酸 I	相距最远	0.3226	0.7	0	0.0331	0.01
15	氨基酸 Q	氨基酸 R	有一个最近共同父类	0.0476	0.8	0.6400	0.6354	0.70

使用欧式距离(  $Distance(X, Y) =$

$\sqrt{\sum_{i=0}^n (X_i - Y_i)^2}$  ) 分别计算公式(3)(刘群<sup>[2]</sup>),公式(1)(Resnik<sup>[3]</sup>),公式(5)(Lin<sup>[5]</sup>)和文中算法(公式(6))与专家期望值之间的差异,结果如表 2 所示:

表 2 各算法与领域专家经验值的欧式距离

序号	算法	欧式距离
1	公式(3)(文献[2])	1.4485
2	公式(1)(文献[3])	2.089
3	公式(5)(文献[5])	0.3753
4	文中算法(公式(6))	0.295

### 3.4 讨 论

从实验结果可以看出,以公式(2)和公式(1)为代表的基于边的算法,忽略了概念节点在本体中的层次信息,例如表 1 中第 1 行和第 8 行数据,虽然语义距离相同,但是第 1 行数据位于本体图的顶层,第 3 行数据位于本体图的底层,第 3 行应当比第 1 行具有更大的相似度。以公式(5)为代表的基于顶点的方法,在一定程度上弥补了基于边的算法的缺陷,但是它忽略了语义距离对相似度的影响,体现在表 1 中第 1 行和第 3 行,虽然它们有相同的最近共同祖先,但是其语义距离不同,因而相似度也应当不同。以上两类算法都忽略了属性等信息,体现在表 1 中第 8 行和第 10 行,虽然其语义距离和最近共同父节点都相同,但是其属性特征差异较大,因此其相似度也不应相同。文中提出的算法(公式(6))将这些因素综合起来考虑,通过实验结果不难发现,其结果比较符合预期。

通过表 2 可以直观的看出,以公式(5)(Lin<sup>[5]</sup>)为代表的基于顶点的算法要优于以公式(3)(刘群<sup>[2]</sup>)和公式(1)(Resnik<sup>[3]</sup>)为代表的基于边的算法。文中算法由于综合考虑了语义距离和信息内容,又加入了属性因子和层次因子的度量,计算结果与专家期望值最小,效果最好。

### 4 结束语

综述所述,文中结合概念在本体图中的层次、距离,共享属性和共享信息等多种特征,提出了一个基于本体结构的相似度算法,并以公开的氨基酸本体为例,证明了算法的有效性。文中的不足之处在于,对领域本体以外的知识考虑较少,在以后的研究中可以借鉴

HowNet, 维基百科<sup>[14]</sup>或者谷歌<sup>[15]</sup>等开源知识库,相信会进一步提高算法精度。

### 参考文献:

[1] Lee W N, Shah N, Sundlass K, et al. Comparison of ontology-based semantic similarity measures [C]//AMIA Annual Symposium Proceedings. USA: [s. n.], 2008:384-388.

[2] 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.

[3] Resnik P. Using Information Content to Evaluate Semantic Similarity [C]//Proceedings of the 14th IJCAI. Montreal, Canada: [s. n.], 1995:448-453.

[4] Chodorow M, Leacock C. Combining local context and WordNet similarity for word sense identification [M]. Cambridge, Massachusetts: MIT Press, 1997:265-283.

[5] Lin D. An information-theoretic definition of similarity [C]//Proceedings of 15th Int. Conf. on Machine Learning. [s. l.]: [s. n.], 1998.

[6] Couto F, Silva M, Coutinho P. Semantic similarity over the gene ontology: Family correlation and selecting disjunctive ancestors [C]//Proc. of the ACM Conference in Information and Knowledge Management (CIKM). [s. l.]: [s. n.], 2005.

[7] Couto F, Silva M. Disjunctive Shared Information between Ontology Concepts: Application to Gene Ontology [J]. Journal of Biomedical Semantics, 2011, 2:2-5.

[8] Tversky A. Features of Similarity [J]. Psychological Review, 1977, 84(4):327-352.

[9] 林智超,朱国进. 一种基于 FCA 的概念相似度算法[J]. 计算机技术与发展, 2008, 18(9):112-114.

[10] 张晓李,王西锋. 基于本体和相似图的概念语义相似度计算[J]. 计算机技术与发展, 2011, 21(8):101-104.

[11] 赵小谦,郑 彦,储海庆. 概念树在短文本语义相似度上的应用[J]. 计算机技术与发展, 2012, 22(6):159-162.

[12] 段寿建,杨朝凤. 基于领域本体的概念语义相似度和相关度综合量化研究[J]. 现代图书情报技术, 2009, 25(11):40-43.

[13] Sabou M, Richards D, van Splunter S. An Experience Report on Using DAML-S [EB/OL]. [2008-12-02]. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.7.529&rep=rep1&type=pdf>.

[14] Cilibrasi R, Vitanyi P. The Google Similarity Distance [J]. Knowledge and Data Engineering, 2007, 19(3):370-383.

[15] Gabrilovich E, Markovitch S. Computing semantic relatedness using Wikipedia-based explicit semantic analysis [C]//IJCAI. [s. l.]: [s. n.], 2007:1606-1611.

(上接第 51 页)

[7] 刘建华. 多色彩高空间分辨率遥感影像矢量边缘信息提取算法与应用[J]. 地球信息科学学报, 2010, 12(6):850-854.

[8] 刘 技,康晓东,贾富仓. 基于图割与均值漂移算法的脊椎

骨自动分割[J]. 计算机应用, 2011, 31(3):760-762.

[9] 辛月兰. 基于图割的图像分割综述[J]. 微型电脑应用, 2012(9):1-6.

# 基于本体结构的语义相似度计算

作者:

杨方颖, 蒋正翔, 张姗姗, [YANG Fang-ying](#), [JIANG Zheng-xiang](#), [ZHANG Shan-shan](#)

作者单位:

[北京理工大学计算机学院, 北京, 100081](#)

刊名:

[计算机技术与发展](#) 

英文刊名:

[Computer Technology and Development](#)

年, 卷(期):

2013, 23(7)

本文链接: [http://d.wanfangdata.com.cn/Periodical\\_wjfz201307013.aspx](http://d.wanfangdata.com.cn/Periodical_wjfz201307013.aspx)