

文章编号:1001-9081(2008)05-1183-03

一种履带式本体构建模型

鲁 强,王智广,陈 明

(中国石油大学(北京) 计算机科学与技术系,北京 102249)

(luqiang@cup.edu.cn)

摘 要:为了有效地指导本体构建,创建了履带式本体构造模型来描述本体的构造方法。定义和描述了构成此模型的基本要素,包括本体(上层本体、领域本体和应用本体)和本体构造步骤(分析、设计、开发、部署和评估),并在此基础上做出了定性分析,说明了顶级本体驱动开发方法、领域本体驱动开发方法和应用领域开发方法各自的特点。通过对此模型的分析,以指导在本体开发中使用正确的开发方法。

关键词:本体工程;本体构建方法;知识工程

中图分类号: TP393 **文献标志码:** A

Track methodology model for ontology building

LU Qiang, WANG Zhi-guang, CHEN Ming

(Department of Computer Science and Technology, China University of Petroleum, Beijing 102249, China)

Abstract: In order to utilize methodology for ontology building more effectively, the Track Methodology Model (TMM) was created. The elements in TMM were defined and described, which included ontology (Top-level Ontology, Domain Ontology and Application Ontology) and five stages in the process of ontology building (analysis, design, development, deployment and evaluation). The qualitative analysis was explored based on the elements. Hence, three methodologies (driving wheel methodology for top-level, domain and application ontology) were imported, while their characteristics and relationships of three methodologies were researched and established. Through the application of building knowledge ontology in university based on TMM, effective methodology for building ontology can be selected to instruct the process of ontology building.

Key words: ontological engineering; ontological methodologies; knowledge engineering

0 引言

本体^[1]是共享概念模型的形式化规范说明,它为不同的异构系统之间交流和通信提供了对领域知识的共同理解,有利于信息系统或智能代理共享和重用知识。对本体的构建是知识工程中关键步骤之一。为了对本体的建设过程实现有效地控制,建立了与软件工程相似的本体工程。本体工程^[2]描述了本体构建过程及其生命周期和各种指导本体构建的方法。其中,文献[3]在 CommonKADS 中将知识模型创建过程分为三个阶段(知识识别、知识规范说明和知识精化)来进行构建,但是 CommonKADS 方法是针对知识建模来进行的,对于本体构建其并不完全适用。在文献[4]方法中,将本体的创建分为五个阶段:分析、设计、开发、部署和评估。各个阶段依次进行,前一个阶段是后一个阶段的基础。这五个阶段循环进行,构成本体开发的螺旋结构。使得本体在维护、演化中不断进行修改,以更好地表达系统中的知识概念。但是,文献[4]方法是针对 P2P 模型中的轻量级本体—应用本体来进行构建的,并没有涵盖和区分其他类型本体的构建。在文献[5]方法中根据本体构造过程中各种因素的影响分析,虽然给出了本体构造时的一些因素的评估方法,但是并没有给出本体构造过程中各个因素之间精确的关系模型。

本文在研究知识和本体构建模型基础之上,建立了履带式构建本体方法模型。在此模型中对要构造的本体按顶级、领域和应用三种类别进行划分,将每种类型本体的构建过程划分为五个阶段:分析、设计、开发、部署和评估。然后,对此模型进行了定性分析,分析和说明了顶级本体驱动开发方法、领域本体驱动开发方法和应用领域开发方法各自的特点及它们之间的关系。最后,通过对高校知识本体构建的应用场景说明了如何根据此模型中的方法和步骤来指导本体的建设。

1 履带式构建本体方法模型

在此模型中,首先描述了要构造的本体类型,说明每种本体作用范围、构造本体的步骤和构造本体步骤之间的关系;然后根据这些本体对此方法模型进行定性分析,给出此模型描述的本体构造过程特性。

1.1 履带式本体构造过程要素

1.1.1 构造的本体

根据文献[6],将构造的本体划分为顶级本体、领域本体(包括任务本体)以及应用本体。其中,上层本体(Top-level Ontology, Tonto)描述的是最普遍的概念及概念之间的关系,如空间、时间、事件、行为等,与具体的应用无关,其他种类的本体都是该类本体的特例;领域本体(Domain Ontology,

收稿日期:2007-11-29;修回日期:2008-02-02。

基金项目:国家自然科学基金资助项目(60072006);北京市重点实验室建设计划项目(JD114140554)。

作者简介:鲁强(1977-),男,河北唐山人,讲师,博士,CCF 会员,主要研究方向:Semantic Web、知识工程、分布式系统;王智广(1964-),男,教授,内蒙古通辽人,CCF 高级会员,主要研究方向:开放式分布计算;陈明(1949-),男,吉林长春人,教授,博士生导师,CCF 高级会员,主要研究方向:开放式分布计算。

Donto)描述的是特定领域(医药、汽车等)中的概念及概念之间的关系;应用本体(Application Ontology, Aonto)描述的是依赖于特定领域和任务的概念及概念之间的关系。

1.1.2 本体构造步骤

根据构造的本体类型,将构造本体的过程分为顶级本体构造过程(Top Ontology Process, TOP)、领域本体构造过程(Domain Ontology Process, DOP)和应用本体构造过程(Application Ontology Process, AOP)。每种类型的本体构造过程中都包括了以下5个步骤,如图1所示。

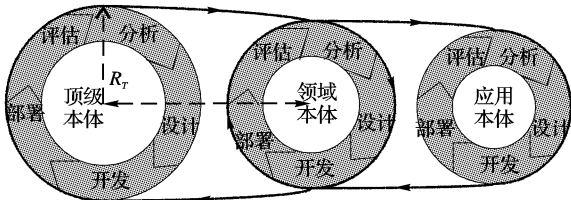


图1 履带式本体构造方法

1)分析阶段是为了从各个方面收集的材料中抽取对建立本体有关的模型,这个模型包括业务内容、同知识相关的问题以及形成执行决定的解决方案等方面的内容。这个阶段参考CommonKADS方法,主要包括两个部分:进行本体范围和可行性研究,识别问题/机会领域和潜在的解决方法,确定构建本体的技术、成本等可行性,以便选择最有希望的构建本体的目标和解决方法;对所选目标解决方案进行分析和改进研究,关注所涉及的任务和本体之间的相互关系、获得成果所需的知识以及可取得的改进等方面内容,并确定组织措施和任务变化,以保证能被组织接受和集成到知识系统解决方案中。

2)设计阶段是将分析阶段所得的模型应用到具体的实际领域中,并确定本体结构、类型以及本体之间的关系。设计阶段包括以下三个步骤:首先从分析模型中划分出解决方案的知识领域及其作用范围;然后在此基础上确定本体概念的内涵和外延以及类型,并创建本体结构以及根据本体之间关系等方面的内容;使用本体建模元语进行表示。设计阶段需要领域专家、知识工程师以及系统用户共同协作来完成。

3)开发阶段是根据在设计阶段形成的本体元语模型,并结合实际的业务领域内容和需求,使用具体的具有使用和开发效率的本体语言来进行本体的表示。这个阶段一般由知识工程师使用特定的本体语言开发工具来完成。在这个阶段完成后,形成可供系统实际使用的本体。

4)部署阶段是在知识系统开发和使用过程中,使得系统开发人员以及系统用户能够熟悉在开发阶段所创建的本体含义、作用范围和使用方法。在这个阶段,设计出来的本体才实际发挥出自己的作用。

5)评估阶段可以发生在上述的任何一个阶段,例如在设计阶段中可以评估设计出来的本体元语模型是否满足分析阶段中所提出的需求模型,在部署阶段评估知识本体的可用性等。但是,最后还需要综合评估创建本体结构的合理性,本体之间的关系是否满足系统要求以及是否存在矛盾,创建的本体粒度是否合理,是否支持未来业务的扩充,是否能够合并一些轻量级本体等方面的内容。

相同的步骤在不同类型的本体上对应不同的内容范围,例如DOP中的分析阶段,是根据顶级本体内容、关系和具体领域的业务内容、规则,来分析领域建立领域本体的有关知识模型,而AOP的分析阶段是根据领域本体和具体的应用内容、业务要求,来分析应用本体的有关知识模型。

在每种类型本体构造过程中这些步骤都是循环反复运行的,以维护和支持对本体的修改、更新。在构建本体过程中使用软件工程中的里程碑方法,每个步骤都有明确的完成目标和完成时间。只有在本次构造本体过程完成后,才能启动下一次本体的构造过程。

1.1.3 不同类型本体构造步骤之间的履带

如图1所示,不同类型本体构造过程的相互影响是通过信息链条来进行连接、传递和驱动的。链条表示不同类型本体构造过程之间的信息传递,即一类本体构造过程对另一类构造过程的影响。启动某种类型本体的构造过程能够通过链条来带动、影响其他类型本体的构造过程。

1.2 模型

设 $Size_i$ 为*i*类型本体的规模大小, $Time_i^j$ 为*i*种类型本体在*j*步骤所消耗的时间,其中*i*表示本体种类(UOnto, DOnto和AOnto);*j*表示本体构建步骤。式(1)表示顶级本体的构建时间是其各个构建步骤所消耗时间的总和。同理可得式(2)、(3)分别表示领域本体和应用本体的构建时间。如图1所示,将 $Time_i$ 表示为*i*类型本体构建步骤形成的圆环周长。

$$Time_{UOnto} = Time_{UOnto}^{Analysis} + Time_{UOnto}^{Design} + Time_{UOnto}^{Develop} + Time_{UOnto}^{Deploy} + Time_{UOnto}^{Evaluate} \quad (1)$$

$$Time_{DOnto} = Time_{DOnto}^{Analysis} + Time_{DOnto}^{Design} + Time_{DOnto}^{Develop} + Time_{DOnto}^{Deploy} + Time_{DOnto}^{Evaluate} \quad (2)$$

$$Time_{AOnto} = Time_{AOnto}^{Analysis} + Time_{AOnto}^{Design} + Time_{AOnto}^{Develop} + Time_{AOnto}^{Deploy} + Time_{AOnto}^{Evaluate} \quad (3)$$

本体的构造过程是循环反复的,如图1所示,因此可以使用本体构建的角速度来描述本体构造过程的平均速率,如式(4)表示*i*类型本体构造过程的角速度。

$$\omega_i = \phi(S, P, M) \quad (4)$$

其中, S 表示本体规模大小, P 表示概念抽象程度变量, M 表示人力、资金因素变量,此函数数值与 S 成反比,与 P, M 成正比。

本体构造过程所构造的本体概念之间的距离,使用本体概念之间相似度的式(5)来表示,其中 R_i 表示*i*类型本体循环构造过程所构造本体的概念平均距离,可以用图1中的构造过程的半径来表示。其中对于本体之间相似度的计算^[7]如式(6)所示,其中 c_1, c_2 表示本体, $lso(c_1, c_2)$ 函数表示最为接近 c_1, c_2 这两个本体概念并且为这两个本体共有的父本体, $p(c)$ 函数表示本体*c*在某个领域出现的概率。

$$R_i = \frac{\sum_{\alpha, \beta \in 1, \alpha \neq \beta} (1 - sim(c_\alpha, c_\beta))}{n \times (n - 1) / 2} \quad (5)$$

$$sim(c_1, c_2) = \frac{2 \times \log p(lso(c_1, c_2))}{\log p(c_1) + \log p(c_2)} \quad (6)$$

将此模型看为封闭的系统,则使用式(7)表示不同本体构造过程之间的相互影响,这些影响是通过不同本体构造过程之间的信息传递来进行的,其中 $\omega \cdot R$ 表示本体构造过程输出的构造信息速率或接受外部本体构造信息的速率。 B_{ij} 为*i*类型本体构造得到向*j*类型本体构造过程传递的信息带宽。

$$\omega_i \cdot R_i = \omega_j \cdot R_j = B_{ij} \quad (7)$$

使用 $D_{\alpha\beta}$ 表示 α 和 β 两种类型本体构造过程中心点之间的距离,用来说明这两种类型本体之间的划分粒度距离,如果两种类型本体之间的划分粒度大,则 D 的数值就大;如果小,则 D 的数值就小,如式(8)所示。其中 n_i, n_j 分别表示*i*和*j*类型本体构造过程中的本体数量。

$$D_{ij} = \frac{\sum_{\alpha \in i, \beta \in j} (1 - sim(c_\alpha, c_\beta))}{(n_i + n_j) \times (n_i + n_j - 1) / 2} \quad (8)$$

T_{ij} 表示为 i 类型本体更新导致 j 类型本体更新的平均时间,如式(9)所示,其中 $i \neq j, \alpha_{ij}$ 表示阻碍 i 向 j 传递的信息影响程度, $0 < \alpha \leq 1$ 。

$$T_{ij} = \frac{\alpha_{ij} \times \sqrt{D_{ij}^2 + (R_i - R_j)^2}}{B_{ij}} \quad (9)$$

1.3 分析

在本模型中按概念的抽象级别将本体划分成顶级本体、领域本体和应用本体,以本体的抽象程度为参数,根据式(4)、(7)可得:

$$R_U > R_D > R_A \quad (10)$$

$$Time_U > Time_D > Time_A \quad (11)$$

即越抽象的概念其构建、维护、更新的周期越长。随着本体构造过程的不断进行,每种类型的本体数量在不断地增加,这时概念与概念之间的距离减小,即 S 变大, R 变小。由式(4)在 P, M 不变的情况下, ω_i 变小, i 类型本体构造时间增加。这是因为随着本体的增多,本体的分析、设计、评估等时间都会增加。由式(7)可知, B_{ij} 变小,即不同类型本体构造过程之间的信息带宽会越来越小。但是根据式(8),在本体数量不断增加的情况下, D_{ij} 逐渐变小,即两个本体构造过程构造的本体概念距离越来越近。

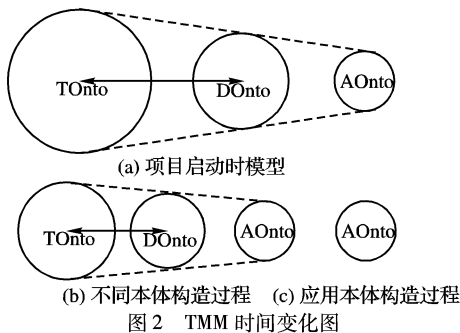


图 2 TMM 时间变化图

根据上面分析,可以知道随着构造过程的循环反复地进行,此模型中的各个本体构造过程发生了变化,如图 2 所示。在项目启动时模型图如图 2(a)所示。随着本体构造时间的增长,不同本体构造过程之间的距离越来越短,构造过程的半径越来越短,导致不同本体构造过程越来越接近,假设构造过程无限进行下去的话,那么此模型将会退回到只有应用本体构造过程,如图 2(c)所示。

2 模型描述的本体构造方法

在图 1 中本体构造过程的时间长短使用代表不同类型本体构造过程的圆周长度来表示,由式(11)可以知道顶级本体创建的时间最长,而应用本体创建的时间最短。每种类型本体的构造过程之间通过链条来连接和驱动。链条表示不同类型本体构造过程之间的信息传递。通过链条连接形象地表明每种类型本体构造的周期频率,顶级本体的更新最慢,领域本体次之,应用本体最快。

把三个构造过程圆环分别看成驱动轮,这样在此模型中描述了三种本体构造方法。它们依次是顶级本体驱动法、领域本体驱动法和应用本体驱动法。顶级本体驱动法是指以顶级本体构造过程作为驱动,首先建立顶级领域本体,在建立顶级本体过程中通过向领域本体传递相关信息来启动领域本体

的构造过程,领域本体构造过程在启动后开始应用本体的构造,将已创建的领域本体信息反馈到顶级本体的构造过程中。应用本体驱动法与顶级本体驱动正好相反。领域本体驱动法是指领域本体构造过程作为驱动,首先建立领域本体,然后同时向顶级本体和应用本体构造过程传递信息,以启动两类本体的构造过程,创建好的顶级本体和应用本体信息反馈到领域本体的构造过程中。

由此可以知道在顶级本体驱动法和应用本体驱动法中信息传递和反馈的方向是单向的,而领域本体驱动法中信息传递和反馈是双向的,即对于单向传递的本体驱动法,完成一次完整的信息传递需要 $T_{ud} + T_{da}$ 时间,而双向驱动法完成时间的传递为 $\max(T_{ud}, T_{da})$ 。三种驱动方法具有不同的应用特点,因此适用于不同的本体开发方案。顶级本体驱动法可以看成是从上到下的螺旋型本体构造方法,它适合于大型的、开发周期长的并且知识构成和范围处于相对稳定的知识系统中的本体构造方法;应用本体驱动法可以看成是从下到上的螺旋型本体构造方法,它适合于开发周期短、知识更新快的小型知识系统中的本体构造方法。在有些小型知识系统中还可以省略顶级本体,甚至领域本体。领域本体驱动法可以看成是以业务来驱动本体开发方法,它处于顶级和应用两者驱动方法之间,具有本体开发周期快、本体更新和维护快等特点,它适用的知识系统类型范围广。但是由于它涉及到将领域本体同时向上抽象和向下细化的两个方向的本体创建过程,因此在实际开发过程中要把握好向上和向下的优先次序,这需要经验丰富的知识工程者来进行控制。

3 应用场景

在构造高校知识本体时应用上述的本体分类方式来进行本体建设,如图 3 所示。在顶级本体层中,包括描述时间、空间、事件、行为、方法等知识本体,这些本体在高校的各个部门和各种知识形式中都会用到,并且具有统一的概念内涵和外延。将领域本体层分为上层领域本体和下层领域本体,在上层领域本体中包括了通讯录本体、组织结构、知识结构本体、文件结构本体等,这些本体描述了高校中各种形式领域的一般特征知识;下层领域本体是对上层领域本体的具体细化,在细化过程中下层领域本体可以引入处在上层的多个领域本体,来表示交叉领域中的知识结构;应用本体层中包括了高校知识系统中具体业务、具体事务以及具体形式的知识本体,这些本体能够直接被知识处理语言利用,用以完成知识系统中的业务逻辑。

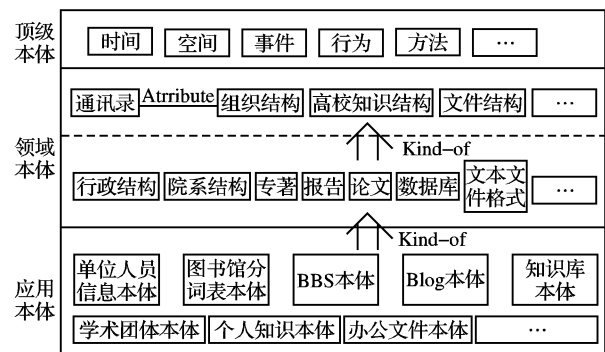


图 3 高校知识本体框架图

在高校本体构建时,由于高校知识结构复杂和种类繁多,为了提高本体的开发和维护速度,选择领域本体驱动法。首
(下转第 1189 页)

- framework for access control: An informal description[C]// Proceedings of the 13th National Computer Security Conference. Washington: NIST Press, 1990: 134 - 143.
- [4] BERTINO E, JAJODIA S, SAMARATI P. Supporting multiple access control policies in database systems[C]// Proceedings of the 1996 IEEE Symposium on Security and Privacy. Washington DC: IEEE Computer Society, 1996: 94 - 107.
- [5] OSBORN S, SANDHU R, MUNAWER Q. Configuring role-based access control to enforce mandatory and discretionary access control policies[J]. ACM Transactions on Information and System Security, 2000, 3(2): 85 - 105.
- [6] Secure Computing Corporation. DTOS lessons learned report, CDRL A008[R]. Roseville, Minnesota: Secure Computing Corporation, 1997.
- [7] BELL D E, LAPADULA L J. Secure computer systems: a mathematical model, ESD-TR-73-278[R]. Bedford, MA: MITRE Corp, 1973.
- [8] HARRISON M H, RUZZO W L, UNMAN J D. Protection in operating systems[J]. Communications of the ACM, 1976, 19(8): 461 - 471.
- [9] BIBA K J. Integrity considerations for secure computer systems, ESD-TR-76-372[R]. Bedford, MA: USAF Electronic Systems Division, 1977.
- [10] DENNING D E. A lattice model of secure information flow[J]. Communications of the ACM, 1976, 19(5): 236 - 242.
- [11] BREWER D F C, NASH M J. The chinese wall security policy [C]// Proceedings of the 1989 IEEE Symposium on Security and Privacy. Washington, DC: IEEE Computer Society Press, 1989: 206 - 214.
- [12] BOEBERT W E, KAIN R Y. A practical alternative to hierarchical integrity policies[C]// Proceedings of 8th National Computing Security Conference. Gaithersburg, MD: [s. n.], 1985: 18 - 27.
- [13] 侯方勇, 周进, 王志英, 等. 可信计算研究[J]. 计算机应用研究, 2004, 21(12): 1 - 4.
- [14] 陈钟, 刘鹏, 刘欣. 可信计算概论[J]. 信息安全与通信保密, 2003, 24(11): 17 - 19.
- [15] 周明辉, 梅宏. 可信计算初探[J]. 计算机科学, 2004, 31(7): 5 - 8.
- [16] 屈延文. 软件行为学[M]. 北京: 电子工业出版社, 2005.
- [17] 林闯, 彭雪梅. 可信网络研究[J]. 计算机学报, 2005, 28(25): 751 - 758.
- [18] 谭良, 刘震, 余堃, 等. CRL 分段—过量发布新模型[J]. 电子学报, 2005, 33(2): 227 - 230.
- [19] 谭良, 余堃, 周明天. CRL 增量—过量发布新模型[J]. 计算机科学, 2005, 32(4): 133 - 136.
- [20] 单智勇. 多安全政策支持框架研究及其在安全操作系统中的实践[D]. 北京: 中国科学院研究生院, 2002.
- [21] 王辉, 刘淑芬. 一种可扩展的内部威胁预测模型[J]. 计算机学报, 2006, 29(8): 1346 - 1355.
- [22] 胡建伟, 汤建龙, 杨绍全. 网络对抗原理[M]. 西安: 西安电子科技大学出版社, 2004.
- [23] 陈爱平. 内部信息行为监管方法[J]. 计算机安全, 2005(8): 53 - 54.
- [24] 谭良, 周明天. 可信操作系统研究[J]. 计算机应用研究, 2007, 24(12): 10 - 15.
- [25] 梁彬. 可信进程机制及相关问题研究[D]. 北京: 中国科学院研究生院, 2004.

(上接第 1185 页)

先启动构建组织结构本体过程,在构建过程中会引用到顶级本体中空间和时间等本体,这时开始启动构建顶级本体过程,来构建相关的顶级本体;在组织结构本体构建过程中还会涉及到业务领域中的人员信息本体,因此来启动构建应用本体过程。在顶级本体和应用本体构造过程中,可以把一些本体构造信息反馈到领域本体构造过程中,以使得要构造的领域本体结构更加合理。通过里程碑的本体开发方法,当与这个业务相关的本体完成全部构建步骤后,启动下一个业务的领域本体构建过程。通过这样反复的根据业务来构造本体的过程,来完成高校知识本体的建设。

4 结语

通过建立本体履带式构造方法模型并对此模型进行定性分析,说明了构造本体的基本步骤和方法,以及每种本体构造方法的特点和适用范围。通过此方法模型,能够在本体构造过程中,实时地评估本体的构造速度、规模和语义距离,可以指导不同类型知识系统的本体开发和构造过程,使得本体开发者能够根据知识系统的特点选择与其相适应的知识本体构造方法。

下一步工作目标是通过已有的大量本体及其构造方法和构造过程中的各种消耗因素进行统计分析,给出函数 ϕ 合理地表示,能够根据要构建的本体规模估计出要构建本体的时间,以及建立不同类型的本体构建之间信息传递速度和信息带宽的计算方法模型。

参考文献:

- [1] STUDER R, BENJAMINS V R, FENSEL D. Knowledge engineering, principles and methods[J]. Data and Knowledge Engineering, 1998, 25(1/2): 161 - 197.
- [2] SOFIA PINTO H, MARTINS J P. Ontologies: How can they be built? [J]. Knowledge and Information System, 2004, 6(4): 441 - 464.
- [3] SCHREIBER G, AKKERMANS H, ANJEWIERDEN A, *et al.* Knowledge engineering and management: The common KADS methodology[M]. Cambridge, MA: MIT Press, 1999.
- [4] MIKA P. HOPE: Harnessing ontologies and peer-to-peer[EB/OL]. [2007 - 10 - 30]. <http://km.aifb.uni-karlsruhe.de/projects/swap/public/public/Publications/swap-d1.2.pdf>.
- [5] SIMPERL E P B, TEMPICH C, SURE Y. ONTOCOM: A cost estimation model for ontology engineering[C]// Proceedings of the 5th International Semantic Web Conference (ISWC 2006), LNCS 4273. Berlin: Springer-Verlag, 2006: 625 - 639.
- [6] GUARINO N. Semantic matching: Formal ontological distinctions for information organization, extraction, and integration[C]// International Summer School on Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology, LNCS 1299. London: Springer-Verlag, 1997: 139 - 170.
- [7] LIN DE-KANG. An information-theoretic definition of similarity [C]// Proceedings of the 15th International Conference on Machine Learning. San Francisco, CA: Morgan Kaufmann Publishers, 1998: 296 - 304.