一种基于异构系统发现日志本体关联规则的方法

孙 明 陈 波 周明天
(电子科技大学计算机科学与工程学院 成都 610051)

摘 要 构建日志本体之上的访问模式关联规则是语义Web使用挖掘的主要任务之一。在DL-safe规则的限定下,将日志本体和一阶应用规则相结合,构成异构日志知识库,以提高Web日志系统的知识表示和推理能力。在此基础上借助ILP理论从异构日志知识库中挖掘出频繁用户访问模式,并生成访问模式关联规则,以发现用户访问行为之间更丰富的潜在关联知识。该方法提高了语义Web使用挖掘的质量,为改进站点结构提供了更有效的决策知识。实验结果证明了该方法的可行性和有效性。

关键词 语义Web使用挖掘,日志本体,异构系统,关联规则,归纳逻辑编程
中图法分类号 TP182 **文献标识码** A

Approach for Discovering Association Rules from Log Ontologies Based on Hybrid System

SUN Ming CHEN Bo ZHOU Ming-tian
(School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610051, China)

Abstract Building access pattern association rules on top of log ontologies is one of the main tasks of semantic Web usage mining. With the restrictions of DL-safe rules, we combined log ontologies with first-order application rules to build a hybrid log knowledge base. It can improve the capability of knowledge representation and reasoning of Web log system. After mining the frequent user-access patterns from the hybrid system through ILP theory, the access pattern association rules can be constructed to discover the potential associations between user-access behaviors. This method improves the results of semantic Web usage mining and provides more decision-making for optimizing the structure of Web sites. The experimental results show that this method is effective and quite feasible to solve practical problems.

Keywords Semantic Web usage mining, Log ontology, Hybrid system, Association rule, Inductive logic programming (ILP)

用户使用信息中的关联知识来反映一个用户访问行为和其它用户访问行为间的依赖或关联。语义Web使用挖掘[1]的主要目的之一就是找出日志系统中隐藏的关联信息,挖掘的结果能够帮助管理者分析相关的设计和商业策略以及构建用户个性化服务,从而获得商业价值的提升。

传统的关联规则发现方法采用Web日志文件作为挖掘集,而本体(Ontology)作为一种在语义层面上的概念建模工具可以表示更丰富的使用背景知识,本体之上的用户访问模式关联规则发现[2]能更清晰地刻画出用户行为的语义关联。但本体以描述逻辑(Description Logic,DL)为基础拥有丰富的概念定义,缺乏对概念和角色之间关系建模的能力,因此日志本体也主要用于描述Web系统中的静态使用知识,难以表达Web日志随站点访问量不断增加而动态变化的用户应用访问语义。一阶Horn逻辑强大的规则推理能力能弥补上述日志本体的不足,为此本文将日志本体和Horn逻辑应用规则相结合,构成异构日志系统,并在此基础上提出一种基于ILP(Inductive Logic Programming,归纳逻辑编程)理论的学习方法。从发现日志本体频繁访问模式出发,构建有效访问模式关联规则,以发现用户访问行为之间的潜在关联,提高语义Web使用挖掘的质量和效率。

1 日志本体

日志本体蕴含了用户访问站点行为的潜在语义,在知识层次上反映了用户和站点交互的行为和兴趣。日志本体采用事件这一核心概念来表述用户访问站点的语义行为。事件(Event)是对一个站点商务模型所有访问任务的详细形式化描述,用集合 E 表示。

根据用户访问目的、访问兴趣和访问策略语义的不同,将事件在共性层次上分为原子事件和复合事件。原子事件(Atom Event,EA)描述用户一次具体的请求,包括内容事件(Content Event,EC)和服务事件(Service Event,ES),分别用于描述用户访问具体页面内容的语义行为和用户使用站点服务的语义行为;而复合事件(Complex Event,EX)是由原子事件构成的一个有序序列,表示语义访问路径和策略。事件之

到稿日期:2009-01-22 返修日期:2009-04-09 本文受国家"十一五"科技支撑计划重大资助项目(2006BAH02A0407),国家自然科学基金(90604033)资助。

孙 明(1978—),男,博士生,主要研究方向为语义Web挖掘以及本体学习等,E-mail:sunm@uestc.edu.cn;陈 波(1977—),男,博士,主要研究方向为粗糙集理论等;周明天(1939—),男,教授,博士生导师,主要研究方向为中间件以及知识发现等。

间主要共性关系包括继承关系 $subClassOf(E_i,E_j)$、整分关系 $hasPart(E_i,E_j)$、前项关系 $previousOf(E_i,E_j)$、实例关系 $instanceOf(E_i,E_j)$以及事件间的其它领域关系。日志本体是描述用户使用信息的一种领域本体,在事件以及事件关系的基础上定义日志本体。

定义 1(日志本体,Log Ontology,LO) 给定使用领域 D,L 是 D 上的逻辑语言。D 上的日志本体定义为一个六元组 $LO:=\{E,\leqslant_E,R,\leqslant_R,F,A\}$。其中,$E$ 和 R 是两个不相交的集合,E 是事件概念集,R 是事件关系集;$\leqslant_E$ 是 E 上的偏序关系,表示事件层次;$\leqslant_R$ 是 R 上的偏序关系,表示关系层次;F 是函数,定义为 $F:E\times E\to R$,表示事件关系;A 为基于 $\mathcal{L}$ 的推导公理集。

图 1 描述了日志本体的主要结构。

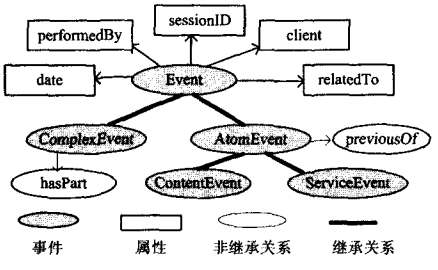

图 1 日志本体主要结构

2 与规则相结合的异构日志知识库

在日志本体的基础上引入表示动态应用语义的一阶规则,构成异构日志知识库,可以大大提高系统的描述能力,但同时也增大了推理上的难度,可能造成系统推理的不可判定性,必须对其结合进行限定。

2.1 DL-safe 规则

与规则相结合的异构系统,如 $\mathcal{AL}$-log[3],CARIN[4] 以及 ORL[5],要么描述逻辑表达能力仍不够丰富,要么推理过程不可判定,在实际应用中存在一定的局限。DL-safe 规则[6]为了消除异构系统推理的不可判定性,对描述逻辑 $\mathcal{SHOIQ}(D)$ 和 Horn 逻辑的子集——选言 Datalog 规则的结合进行限定,确保可判定的同时可以引入更丰富的描述逻辑。

定义 2(DL-safe 规则[6]) 设 KB 为一个 $\mathcal{SHOIN}(D)$ 知识库,N_C 为知识库中的概念集合,N_{Ra} 为抽象角色的名字集合,N_{Rc} 为类型角色的集合,N_P 为谓词符号的集合,$\{\approx\}\cup N_C\cup N_{Ra}\cup N_{Rc}\subseteq N_P$。一个 DL 项的形式如 $A(s)$,其中 $A\in N_C$,或者形如 $R(s,t)$,其中 $R\in N_{Ra}\cup N_{Rc}$。规则 r 形式为

$$A_1\vee A_2\vee\cdots\vee A_m\leftarrow B_1,B_2,\cdots,B_n \tag{1}$$

规则 r 的前件中只能包含逻辑合取运算符,谓词符号来源于 N_P 集合与非 DL 项谓词,并且所有在 r 的前件中的变量必须至少出现在一个非 DL 项的谓词中,这是为了确保所有被规则使用的个体都是被显式引入到知识库中的。

这样定义的结果增加了描述逻辑与规则的表达力,同时确保了系统的可判定性。相应的算法也在 KAON2 中实现。在实际推理过程中,可以通过为规则添加非 DL 原子 $\mathcal{O}(x)$并在知识库中为每一个体 a 添加基原子 $\mathcal{O}(a)$的方式来满足 DL-safe 限制。

2.2 结合 DL-safe 规则的异构日志知识库

DL-safe 规则对描述逻辑与 Datalog 子句规则的结合进行了较为强势的限定,其最大的问题在于不能在推理中基化匿名个体,但是日志本体中事件的个体都是有名个体,从而避免了上述问题的出现。接下来定义基于 DL-safe$^{\mathcal{L}}$ 的异构日志知识库,DL-safe$^{\mathcal{L}}$ 是在描述逻辑语言 L(这里 L 限定为 $\mathcal{SHIQ}(D)$)上满足 DL-safe 规则的一组选言 Datalog 语言。

定义 3 DL-safe$^{\mathcal{L}}$ 知识库 $B=(KB,P)$,其中 KB 是对应于描述逻辑 L 的日志本体知识库,P 是满足 DL-safe 限制的应用规则 r 组成的集合,r 形如

$$h(X)\leftarrow a_1(Y_1),\cdots,a_m(Y_m)\&b_1(Z_1),\cdots,b_n(Z_n) \tag{2}$$

其中,$h,a_1,\cdots,a_m$ 是非 DL 谓词,$b_1,\cdots,b_n$ 是 DL 谓词,$X,Y_1,\cdots,Y_m$ 是任意元项的序列,$Z_1,\cdots,Z_n$ 是一元或二元项的序列。

为了在统一的规则下挖掘关联知识,需要将基于描述逻辑语言 L 知识库 KB 等价地转换为一阶选言 Datalog 规则库 $DD(KB)$,并在此基础上附加 DL-safe 程序 P,构成新知识库 $DD(KB)\cup P$,参考文献[7]给出的具体转换过程。为了描述方便,后文 DL-safe 规则均省略特殊非 DL 谓词 $\mathcal{O}$。

3 基于异构系统的关联规则发现

用户访问模式描述了用户在特定访问策略下使用站点的行为,而访问模式关联规则蕴含了用户行为之间的关联知识。用户行为之间的关联是复杂的,蕴藏在 Web 使用数据中。关联知识挖掘的目的就是通过关联分析找出这些隐藏的关联信息,因而对商业决策具有新的价值。

3.1 访问模式的关联规则

在 DL-safe$^{\mathcal{L}}$ 异构日志知识库 B 上给定基准事件 E_{ref},用户访问模式 H 是 E_{ref} 和任务相关事件间构成的一组联合查询,称为 ε-查询,其应答集由 E_{ref} 的个体组成。ε-查询 H 是具有以下形式的 DL-safe 规则

$$H=q(X)\leftarrow a_1(Y_1),\cdots,a_m(Y_m)\&E_{ref}(X),b_1(Z_1),\cdots,b_n(Z_n) \tag{3}$$

其中,$a_1,\cdots,a_m$ 是 Datalog 谓词,$b_1,\cdots,b_n$ 是 DL 谓词,X 是特征变量。特别地,$H_t=q(X)\leftarrow\&E_{ref}(X)$称为平凡 ε-查询,这里 H_t 子句体中省略了特殊非 DL 原子 $\mathcal{O}(x)$。查询 H 的应答是一个基置换 θ,若 $B\vdash H\theta$,则称 θ 是查询 H 关于知识库 B 的正确应答。应答集 $answerset(H,E_{ref},B)$是 H 关于 B 的正确应答构成的集合。

访问模式关联规则的产生基于 DL-safe$^{\mathcal{L}}$ 上的偏序关系。本文基于 B-包容[8]描述这种偏序关系:给定 DL-safe$^{\mathcal{L}}$ 上的两个访问模式 H_1,H_2,它们之间的偏序关系记为 $H_1\geq_B H_2$。同时,若 H_2 是在 H_1 的子句体上通过添加新的原子而构成的,则 H_1 子句体原子构成的集合是 H_2 子句体原子构成集合的子集,记为 $H_2\supset H_1$。

定理 1 给定 DL-safe$^{\mathcal{L}}$ 知识库 $B,H_1,H_2\in$ DL-safe$^{\mathcal{L}}$ 是 B 上的访问模式,若 $H_2\supset H_1$,则有 $H_1\geq_B H_2$。

证明:$H_2\supset H_1$ 表示 H_2 是在 H_1 的子句体上添加原子 T 而构成的。文献[9]已证明制约 Datalog 程序上,若 $H_1\geq_B H_2$,当且仅当存在一个 H_1 的置换 θ 使得 $head(H_1)\theta=head(H_2)$且 $B\cup body(H_2)\sigma\vdash body(H_1)\theta\sigma$,这里 σ 是 Skolem 置换,而 $body(H_1)\theta\sigma$ 是基 Datalog 子句。

令置换 θ 为空置换,σ 是 B 上的 Skolem 置换。显然 $head(H_1)\theta=head(H_2)$,且 $body(H_2)\sigma=body(H_1)\sigma\cup T\sigma$。因此

$B \cup body(H_2)\sigma \vdash body(H_1)\theta\sigma$。既然 σ 是一个 Skolem 置换，$body(H_1)\theta\sigma$ 就是基 Datalog 子句，因此 $H_1 \geq_B H_2$。证明完毕。

访问模式关联规则描述了一个用户访问模式与其它访问模式之间的关联度，在偏序关系上体现为模式 H_1 加添原子构成模式 H_2 之后 H_1 推导出 H_2 的可能性。作为模式扩展，在访问模式偏序关系的基础上定义关联规则如下。

定义 4 给定模式 $P,Q \in$ DL-safe$^{\mathscr{S}}$，且 $P \supset Q$，则 DL-safe$^{\mathscr{S}}$ 上访问模式关联规则具有以下形式：$Q \Rightarrow (P/Q)(s,c)$，其中 s 和 c 分别代表关联规则的支持度和可信度。

3.2 关联规则挖掘任务

发现异构日志知识库中访问行为关联规则是指在特定访问策略下找出用户使用站点频繁行为之间的关联。该任务定义如下：在 DL-safe$^{\mathscr{S}}$ 异构日志知识库 B 和观察集 $O(B \cap O=\emptyset)$ 之上，给定基准事件 $E_{ref} \in EX$、最小支持度 $minsup$ 和最小可信度 $minconf$，从基准事件 E_{ref} 出发找出所有频繁模式 H，并构建满足 $s \geq minsup$ 且 $c \geq minconf$ 的关联规则 R。

由于复合事件在更高的层次上表示了用户的访问策略，因此基准事件 E_{ref} 限定为复合事件。接下来定义访问模式的支持度，并在此基础上给出访问模式关联规则的支持度和可信度的定义。

定义 5 给定模式 $H \in$ DL-safe$^{\mathscr{S}}$，其支持度定义为 H 与其平凡 ϵ-查询 H_{ref} 应答集中基准事件个体数之间的百分比

$$support(H,E_{ref},B)=\frac{|answerset(H,E_{ref},B)|}{|answerset(H_{ref},E_{ref},B)|} \quad (4)$$

定义 6 给定 DL-safe$^{\mathscr{S}}$ 关联规则 $Q \Rightarrow (P/Q)(s,c)$，其支持度 s 定义为模式 P 的支持度，可信度 c 定义为模式 P 和 Q 的支持度之比

$$s=support(P,E_{ref},B) \quad (5)$$

$$c=support(P,E_{ref},B)/support(Q,E_{ref},B) \quad (6)$$

3.3 发现关联规则的 ILP 方法

从异构日志知识库中发现访问模式关联规则的学习方法首先是发现访问模式中的频繁模式集，然后根据关联规则的定义找出满足 $s \geq minsup$ 和 $c \geq minconf$ 的有效访问模式关联规则。下面详细讨论发现访问规则关联模式的方法。

采用类似文献[10]的方法在异构日志知识库 B 上建立用户访问模式集 CP，其中 CP 是一棵具有层次结构的模式树，在此基础上借助 ILP 理论对模式进行实例检测并计算其支持度 s。若 $s \geq minsup$，则该模式为频繁模式，所有频繁模式构成频繁模式集 FP。在 ILP 的框架下对访问模式 H 进行实例检测，要求 H 与知识库 B 逻辑蕴含观察集 $O(B \cup H \vdash O)$。具体来说就是对于基准事件 E_{ref} 的个体 a_i 要求 $B \cup H \vdash O_i$，其中 $o_i \in O$ 定义为一个二元组 $(q(a_i),A_i)$ 表示 a_i 对应的观察，$q(a_i)$ 是基原子，A_i 是一组关于 a_i 的基 Datalog 事实。接下来给出 H 关于 B 蕴含 o_i 的定义。

定义 7 给定模式 $H \in$ DL-safe$^{\mathscr{S}}$、知识库 B 以及观察 $o_i \in O$，H 关于 B 蕴含 o_i 当且仅当 $B \cup H \cup A_i \vdash q(a_i)$。

引理 1 Q 为 DL-safe$^{\mathscr{S}}$ 知识库 B 的基查询，若 $B \vdash Q$ 当且仅当 $B \models Q$。

证明过程与文献[8]类似，这里不再赘述。

定理 2 给定模式 $H \in$ DL-safe$^{\mathscr{S}}$、知识库 B 以及观察 $o_i \in O$，若 H 关于 B 蕴含 o_i 当且仅当 $B \cup H \cup A_i \vdash q(a_i)$。

证明：P 关于 B 蕴含 o_i

$\Leftrightarrow B \cup H \cup A_i \models q(q_i)$ （定义 7）

$\Leftrightarrow B \cup P \cup body(o_i) \vdash q(a_i)$ （引理 1）

因此 E_{ref} 的个体 a_i 关于 H 的实例检测就等价于是否能为查询 $\leftarrow q(a_i)$ 找到关于 $B \cup H \cup A_i$ 的正确应答。由于异构日志知识库中为了获得更为实际和灵活的访问模式，在模式构建的过程中引入了一阶应用规则集 P 中的非 DL 原子，因此对 E_{ref} 个体 a_i 关于 H 的实例检测又应当分为两个阶段进行：首先采用制约 SLD 否证的方法消除模式中非 DL 原子，然后将基化后的 DL 原子提交至描述逻辑推理机完成实例检查。参考 $\mathscr{AL}$-log[4]，定义 DL-safe$^{\mathscr{S}}$ 上的制约 SLD 否证。

定义 8（制约 SLD 否证） 给定 DL 语言 L，$B=(KB,P)$ 是 DL-safe$^{\mathscr{S}}$ 知识库。q 是关于 B 的查询目标，其制约 SLD 否证是一组有限的制约 SLD 演算 $d_1,\cdots,d_m$ 满足下面条件，其中 $q_0^i,\cdots,q_m^i$ 是第 i 个演算 d_i 的查询目标序列，$1 \leq i \leq m$。

(1) 对于任意 i，q_m^i 形如 $\&\beta_1,\cdots,\beta_k$，即每个演算 d_i 的最后查询目标 q_m^i 是制约空子句；

(2) 对于任意 B 的模型 Γ，至少存在某个 i，$1 \leq i \leq m$，使得 $\Gamma \models q_m^i$，记作 $B \models disj(q_{m1}^1,\cdots,q_{mn}^n)$。

如果关于 B 的查询目标 q 有制约 SLD 否证，那么记作 $B \vdash q$。经过制约 SLD 否证后所有 DL 原子得以全部基化，然后将 DL 基原子提交 DL 推理机完成实例检测，找出所有满足条件的事件个体构成模式的应答集，并根据定义 5 计算出模式的支持度。所有满足大于最小支持度 $minsup$ 的模式构成频繁模式集 FP。

对于每一个频繁模式 $P \in FP$ 找出其所有父模式 $Q \in FP$ $(P \supset Q)$，构建关联规则 $R=Q \Rightarrow (P/Q)(s,c)$ 并计算 R 的可信度 c。若满足 $c \geq minconf$，则将其加入关联规则集 AR（由定义 6 知，关联规则 R 的支持度定义为模式 P 的支持度，因为 P 已经是频繁模式，故 R 的支持度一定满足 $s \geq minsup$）。基于异构日志知识库发现访问模式关联规则算法如图 2 所示。

```
输入：异构日志本体库 B，基准事件 E_ref
初始化 FP=∅，AR=∅；
基于核心事件 E_ref，在 B 上构建访问模式集；
foreach H∈CP do
    基于制约 SLD-否证消解 H 中的非 DL 原子；
    将 H 提交至 DL 推理机进行实例检测并计算其支持度 s；
    if s≥minsup then
        将 H 计入频繁模式集 FP=FP∪H；
    endif
endfor
foreach H∈FP do
    计算 H 所有的父模式，构成父模式集 HP；
    foreach Q∈HP do
        计算可信度 c=support(H,E_ref,B)/support(Q,E_ref,B)；
        if s≥minconf then
            构建关联规则 Q⇒(P/Q)(s,c)；
            将关联规则加入规则集 AR，
            AR=AR∪{Q⇒(P/Q)(s,c)}；
        endif
    endfor
endfor
结束；
输出 AR
```

图 2 访问模式关联规则发现算法

4 仿真实验

为验证本文提出方法的有效性和可行性，以 Java 为编程语言实现方法原型并进行仿真实验。测试环境采用 Intel Core2 2.2G，2G 内存，Windows 2003 server sp2，J2SDK1.5，推理机为 KAON2。测试数据集 OMS 是利用 OntoManager[11] 和 Protégé 根据某手机电子商务网站 8 个月的日志文件生成的日志本体结合应用规则构成的异构日志知识库，包括 58 个事件（包括 4 个复合事件）和 18 个角色，27723 个实例，应用规则 11 条，观察事实 923 个。部分异构日志知识库内容形式化描述如表 1 所列。

表 1 部分异构日志知识库内容形式化描述

日志本体	内涵	Shopping⊆EX Pay⊆ES MobilePhone⊆EC MobilePhone⊆Product Samsung⊆MobilePhone Shopping∩∃hasPart. Pay⊆BuyProduct BuyProduct∩BrowseProduct=∅
	外延	Shopping('ex0806012') Pay('espay0806012') Samsug('E908') hasPart('ex0806012','espay0806012') hasPart('ex0806012','E908')
DL-safe 规则		DicretBuy(x)← searchService(x,3),𝒪(y) &Shopping(x),hasPart(x,y),Pay(y)

观察集包括非 DL 观察事实，如 *searchTimes*（'*ex*0806012',3)以及 𝒪('*espay*0806012')等。在上述测试集中进行访问模式关联规则的测试，给定基准事件 *BuyProduct* 以及 *minsup*=0.2，*minconf*=0.6。在异构日志知识库上访问模式集 *CP* 和频繁模式集 *FP* 均采用树形结构构成模式之间的继承关系。测试结果如表 2 所列。

表 2 OMS 上访问模式关联规则测试结果

FP 层次	频繁模式数量	访问模式关联规则数量	运行时间(秒)
2	5	2	27.8
3	16	14	130.2
4	57	61	904
5	149	163	9061.8
6	255	312	26110.1

测试结果显示了在异构日志知识库不同 *FP* 层次上生成频繁模式和关联规则的数量以及运行时间。发现的访问模式关联规则描述了丰富的语义关联知识，例如在 *FP* 第 3 层以及第 4 层分别发现频繁模式

$$q(x)\leftarrow DicretBuy(x)\&BuyProduct(x)hasPart(x,y) \tag{7}$$

$$q(x)\leftarrow DicretBuy(x)\&BuyProduct(x)hasPart(x,y),\ Samsug(y) \tag{8}$$

它们之间构成的访问模式关联规则满足两个最小阈值

$$DicretBuy(x)\&BuyProduct(x)hasPart(x,y)\Rightarrow Samsug(y)(0.23,0.67) \tag{9}$$

表示用户购物行为中直接购物策略有可能购买的物品是三星产品，该关联规则的支持度和可信度分别为 0.23 和 0.67。

从性能测试可以看出本方法的计算复杂度为 EXPTIME，这是因为异构日志知识库的描述逻辑和一阶规则分别为 $\mathcal{SHIQ}(D)$ 以及选言 Datalog 程序，$\mathcal{SHIQ}(D)$ 中实例检测的计算复杂度为 EXPTIME-完全，基于选言 Datalog 的 DL-safe 程序复杂度却是 EXPTIME，而本文提出的访问模式关联规则挖掘算法是在将模式提交到异构日志知识库进行实例测试前利用制约 SLD 否证的方法对其进行推理消除 Datalog 原子，可以看成是两种逻辑上推理算法的直接连接，因此其计算复杂度仍限定在 EXPTIME 范畴内。

结束语 本文在 DL-safe 规则的限定下将以描述逻辑为基础的日志本体和以一阶 Horn 逻辑为基础的应用规则相结合，构成异构日志知识库，提高了 Web 日志系统的语义表示和推理能力。在此基础上借助 ILP 理论发现频繁用户访问模式并构建有效访问模式关联规则，发现了用户访问站点行为之间的潜在关联知识，提高了语义 Web 使用挖掘的质量，使得挖掘结果更为准确和丰富，为改进站点结构提供个性化服务提供了丰富的决策知识。

参考文献

[1] Berendt B, Hotho A, Stumme G. Usage. Mining for and on the Semantic Web[C]// Data Mining Next Generation Challenges and Future Directions. Boston: AAAI Press, 2004: 461-481

[2] Józefowska J, Ławrynowicz A, Łukaszewski T. On reducing redundancy in mining relational association rules from the Semantic Web[C]// Proceedings of the 2nd International Conference on Web Reasoning and Rule Systems, RR 2008. Heidelberg, Germany: Springer Verlag, 2008: 205-213

[3] Motik B, Sattler U, Studer R. Query answering for OWL-DL with rules[J]. Web Semantics: Science, Services and Agents on the World Wide Web, 2005, 3(1): 41-60

[4] Francesco M D, Maurizio L, Daniele N, et al. AL-log: Intergrating datalog and description logics[J]. Intelligent Information Systems, 1998, 10(3): 227-252

[5] Levy Y A, Rousset M C. Combining Horn rules and description logics in CARIN[J]. Artifician Intelligence, 1998, 104(1/2): 165-209

[6] Horrocks I. A proposal for an OWL rules language[C]// Proceedings of the Thirteenth International World Wide Web Conference. New York: ACM, 2004: 723-731

[7] Motik B, Sattler U. A comparison of reasoning techniques for querying large description logic ABoxes[C]// Proceedings of the 13th International Conference on Logic for Programming Artificial Intelligence and Reasoning, LPAR 2006. Heidelberg: Springer Berlin, 2006: 227-241

[8] Francesca A L. An ILP approach to ontology refinement for the semantic Web[C]// Proceedings of the 4th European Semantic Web Conference. Springer, 2007

[9] Francesca A L, Donato M. Inducing multi-level association rules from multiple relations[J]. Machine Learning, 2004, 55: 175-210

[10] Jozefowska J, Lawrynowicz A, Lukaszewski T. A study of the SEMINTEC approach to frequent pattern mining[C]// Proceedings of PriCKL 2007, ECML/PKDD'2007 Workshop on Prior Conceptual Knowledge in Machine Learning and Knowledge Discovery. Warsaw, Poland. 2007: 41-52

[11] Nenad S, Jorge G, Ljiljana S. ONTOLOGER: A system for usage-driven management of ontology-based information portals [C]// Proceedings of the 2nd International Conference on Knowledge Capture, K-CAP 2003. New York: ACM, 2003: 172-179