# 基于 IF 理论的分布式系统语义互操作研究*)

鲍 泓[1] 封军康[2] 刘宏哲[1]

(北京联合大学信息技术研究所 北京 100101)[1] (英国 paisley 大学计算机学院)[2]

**摘 要** 信息流理论是关于分布式系统中信息流动的数学模型,它关注信息中的语义,可作为语义信息理论的基础。本文主要介绍在异构的分布式数字博物馆语义互操作的研究和应用中,信息流作为一种分析和实现语义互操作的理论框架,为研究分布式系统的语义互操作问题提供了理论基础。

**关键词** 信息,信息流,分布式系统,语义互操作

## IF Theory-based Semantic Interoperability for Distributed Information System

BAO Hong[1] FENG Jun-Kang[2] LIU Hong-Zhe[1]

(Institute of Information Technology, Beijing Union University, Beijing 100101)[1]

(School of Computing, University of Paisley, Paisley, PA1 2BE, UK)[2]

**Abstract** Information flow is a mathematical model of information flow, which focuses on semantics in information and may be seen as a foundation for semantic information theory. Show that the theory to be applicable in semantic interoperability for heterogeneous distributed digital museums.

**Keywords** Information, Information flow, Distributed system, Semantic interoperability

## 1 概述

有关信息的概念随处可见:信息管理、信息系统、信息集成等等。我们正处于"信息时代"以及由信息所引发的革命浪潮中。

目前有关于信息的变革主要是基于技术层面的,人们不断地探索新的以及更有效的信息转换和传输的手段。技术革新应以相应的理论为指引,但是目前尚没有一套完善的关于信息科学的理论。长期以来支持我们进行信息交流的理论基础是 Shannon 的信息论,它是关于通信的数学理论。在当前网络信息时代,人们关注的不仅是信息传输,更重要的是信息中的语义。Dreske,Devlin 和 Floridi 等人的工作都是试图建立这样一套理论[1~4],在前人工作的基础上,Barwise 和 Seligman 提出了"通道理论(channel theory)",二人于 1997 年的专著中提出了两个关于分布式系统中信息流动的数学模型:信息通道(information channel)和局部逻辑(local logies)[5]。我们遵循 Kent,Kalfoglou 和 Schorlemmer 的说法,称此理论为信息流(information flow)理论,简称 IF 理论。

IF 理论是关于存在某种特定联系的一组对象间信息如何流动的复杂数学模型。对象间的这种特定的联系蕴含着一定的规则(regularity),根据这些规则将对象集合模型化为分布式系统。也就是说,此理论是关于分布式系统中信息流动的理论。此处的分布式系统不一定是物理上存在的,也可以是概念上的[6]。它是目前较成熟的语义信息理论之一,可作为研究语义互操作的理论基础。

## 2 相关信息流理论介绍

### 2.1 信息流理论的四项原则

Barwise 和 Seligman 的 IF 理论是基于以下四项原则:

- 信息流动源于分布式系统中的规则;
- 信息流动取决于类型(type)和实例(token);
- 分布式系统中联系规则蕴含着系统的一些组成部分,承载着另一些组成部分的信息;
- 在构造信息通道的分析过程中,一个给定的分布式系统中的规则是相对的。

### 2.2 分类(classification)

分类可以表示为结构 $A = \langle \mathrm{tok}(A), \mathrm{typ}(A), \models_A \rangle$,tok(A)是指待分类的对象集合,称为 A 的实例;typ(A)是指用来划分实例的对象集合,称为 A 的类型。$\models_A$ 指实例和类型间分类关系的二元关系。如果 $a \models_A \alpha$,我们称在分类 A 中实例 a 是属于类型 $\alpha$ 的。通常我们用图 1 表示分类关系。

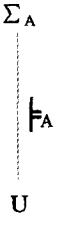


图 1 分类(classification)

### 2.3 信息射(infomorphisms)

如果 $A = \langle \mathrm{tok}(A), \mathrm{typ}(A), \models_A \rangle$,$C = \langle \mathrm{tok}(C), \mathrm{typ}(C), \models_C \rangle$为两个分类,对于分类 C 的所有实例 c 以及分类 A 的所有类型 $\alpha$,满足 $f^\vee(c) \models_A \alpha$ iff $c \models_C f^\wedge(\alpha)$,则称 A 与 C 之间的互逆函数对 $f = (f^\wedge, f^\vee)$为 A、C 之间一个信息射($f^\wedge$读作"f-up",$f^\vee$读作"f-down"),记作 $f: A \rightleftarrows C$,如图 2 所

*)北京市教育委员会科技发展计划项目(km200611417001);鲍 泓 教授,主要从事分布式系统、软件工程研究;封军康 英国 PAISLFY 大学高级讲师,研究语义信息理论。

示。

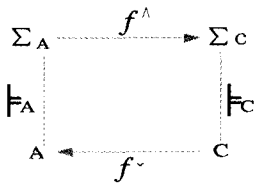


图 2 信息射(infomorphisms)

例如,存在两个分类 Punct 和 Sent,Punct 的实例是指文字中出现的一个个标点符号,例如逗号、句号、问号等。我们将 Punct 按逗号、句号、问号进行分类,此处出现的逗号、句号、问号是指分类类型。Sent 的实例是指文章出现的一个个语句,Sent 的类型是指陈述句、问句及其它。

我们定义 Punct 和 Sent 之间的信息射 $f$: $Punct \rightleftarrows Sent$ 如下:在实例层面上,$f^{\vee}$ 从右至左将出现在文章中的一个个带标点符号的语句转化为对应的标点符号;在类型层面上,$f^{\wedge}$ 从左至右地将 Punct 中的句号和逗号类型转换为 Sent 中的陈述句类型,将问号类型转化为问句类型。

信息射使我们能够构造信息通道(information channel),它体现分布式系统的某一个组成部分是如何与整个系统联系起来的。

### 2.4 信息通道(information channel)

如图 3 所示,信息通道表示为 $C=\{f_i: A_i \rightleftarrows C\}_{i\in I}$ 的信息射族组成,它们具有共同的域 $C$,成为此信息通道的核(core)。$A_i$ 是系统 $C$ 的各个组成部分,正是因为 $A_i$ 是 $C$ 的组成部分,所以各个 $A_i$ 承载着相互之间的信息。

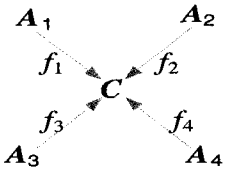


图 3 信息通道

信息通道基本架构如图 4 所示。

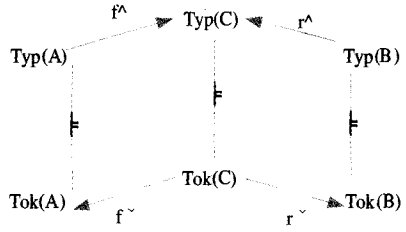


图 4 二元信息通道

### 2.5 信息流动(information flow)

Barwise 和 Seligman 提出关于信息如何流动的定义为:给定 $A$ 和 $B$ 是具有核 $C$ 的信息通道的组成部分的分类。如果实例 $a$ 和 $b$ 由核 $C$ 连接,那么属于类型 $\alpha$ 的实例 $a$ 承载着属于类型 $\beta$ 的实例 $b$ 的信息,在核 $C$ 的理论 Th($C$)中 $\alpha$ 的转变意味着 $\beta$ 的转变。

### 2.6 规则理论

理论(theory)是为了说明系统中究竟什么将在分布式系统中流动,它由若干个约束(constraints)构成。

给定一个分类 $A$,一个相继(sequent)是由 $A$ 的类型 Typ($A$)组成的序偶($\Gamma$, $\Delta$)。如果 $A$ 的实例 $a$ 满足以下条件,则称 $a$ 满足相继($\Gamma$, $\Delta$)。

$(\forall \alpha \in \Gamma)[a \vdash \alpha] \Rightarrow (\exists \alpha \in \Delta)[a \vdash \alpha]$

如果分类 $A$ 中的每个实例都满足相继($\Gamma$, $\Delta$),则称在分类 $A$ 中 $\Gamma$ 蕴涵(entails)$\Delta$,记作 $\Gamma \vdash_A \Delta$。如果 $\Gamma \vdash_A \Delta$ 成立,则称相继($\Gamma$, $\Delta$)为分类 $A$ 中的一个约束(constraints)。

分类 $A$ 所有约束的集合称作 $A$ 的完整理论(complete theory),记作 Th ($A$)。分类 $A$ 的完整理论是分类 $A$ 在系统中的规则的表示。

### 2.7 局部逻辑(Local Logics)

一个局部逻辑 $L = \langle A, \vdash_L, N_L \rangle$ 包括分类 $A$,分类 $A$ 的相继集合 $\vdash_L$,称为 $L$ 的约束(constraints),以及满足 $L$ 的约束 $\vdash_L$ 的那部分实例子集 $N_L \subseteq A$,称为 $L$ 的平常实例(Normal tokens)。

一个局部逻辑 $L$ 的可靠性(soundness)和完备性(completeness)是如下定义的:如果 $L$ 的每一个实例都是正常的(normal),则它是可靠的。如果所有的正常实例所蕴含的相继都包含在 $\vdash_L$ 中,那么称 $L$ 是完备的(complete)。

## 3 IF 在语义互操作中的应用

信息流理论已经逐渐在国际上得到重视,并有不少应用。本文着重介绍它在语义集成中的应用。Robert Kent 是最早应用信息流理论的学者,他说明了本体共享如何在信息流的概念知识模型中能够被形式化[7,8]。Schorlemmer 和 Kalfoglou 在 Robert Kent 的研究基础上进一步讨论了信息流支持语义互操作的问题[9,10],他们指出,信息流可以作为一种实现和分析语义互操作的理论框架。

在以上研究的基础上,我们将信息流理论应用于异构数字博物馆间的语义互操作上[11,12]。在众多语义不一致中,博物馆间对文物分类标准不一致是一个比较突出的问题。例如,在对青铜器的分类中,有的按照时期进行分类,有的是按其功用进行分类,导致系统间相互查询时的语义障碍。


图 5 分期铜器

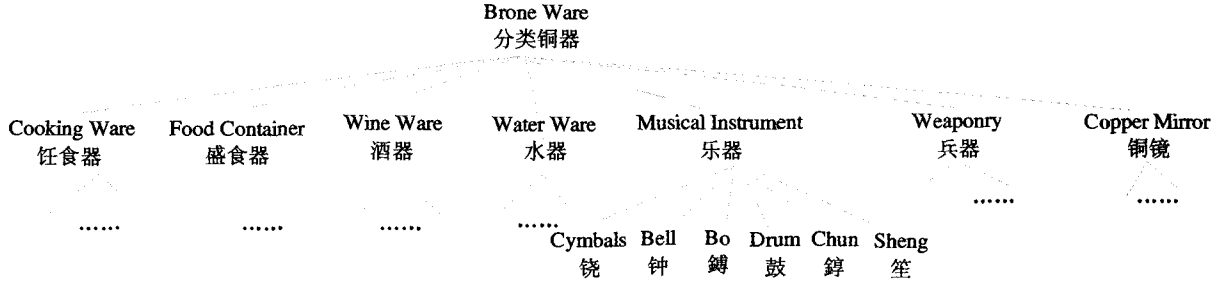


图 6 分类铜器

为解决以上问题，我们利用了信息流的二元信息通道，通过两对互逆函数表明系统间知识共享由类型和实例(Normal Token)两个层面来决定，实例的连接关系对其同样起着至关重要的作用。将两个数字博物馆的分类分别看成是 Classification A，Classification B，而 Classification C 则是连接 Classification A 和 Classification B 的一致分类，如图 7 所示。

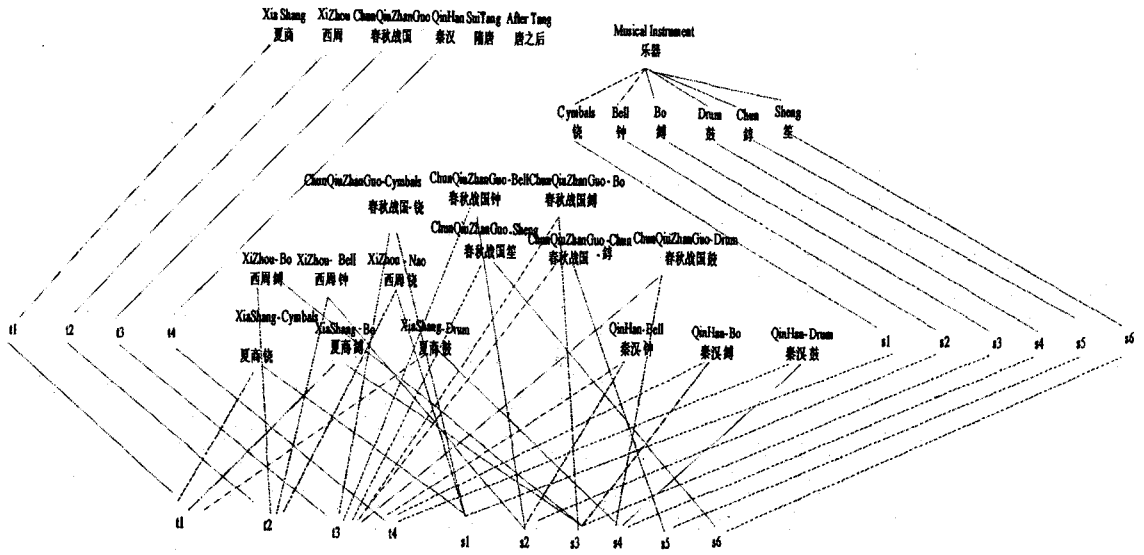


图 7　铜器分类信息通道

Classification A:

Type: XiaShang, XiZhou, ChunQiuZhanGuo, QinHan

Token: $t1$, $t2$, $t3$, $t4$

Classification B:

Type: Cymbals, Bell, Bo, Drum, Chun, Sheng

Token: $s1$, $s2$, $s3$, $s4$, $s5$, $s6$

Classification C:

Type: XiaShang-Cymbals, XiaShang-Bo, XiaShang-Drum, XiZhou-Cymbals, XiZhou-Bell, XiZhou-Bo, ChunQiuZhanGuo- Cymbals, ChunQiuZhanGuo-Drum, ChunQiuZhanGuo-Bell, ChunQiuZhanGuo-Bo, ChunQiuZhanGuo-Nao, ChunQiuZhanGuo-Chun, QinHan-Cymbals, QinHan - Bell, QinHan - Bo

Token: $\langle t1,s1\rangle$, $\langle t1,s3\rangle$, $\langle t1,s4\rangle$, $\langle t2,s1\rangle$, $\langle t2,s2\rangle$, $\langle t2,s3\rangle$, $\langle t3,s1\rangle$, $\langle t3,s2\rangle$, $\langle t3,s3\rangle$, $\langle t3,s4\rangle$, $\langle t3,s5\rangle$, $\langle t3,s6\rangle$, $\langle t4,s2\rangle$, $\langle t4,s3\rangle$, $\langle t4,s4\rangle$

图 7 表明两个不同的分类标准将互通机制建立在一致的分类标准上。

表 1 将实例连接对 $\langle t1,s1\rangle$, $\langle t1,s3\rangle$, $\langle t1,s4\rangle$, $\langle t2,s1\rangle$, $\langle t2,s2\rangle$, $\langle t2,s3\rangle$, $\langle t3,s1\rangle$, $\langle t3,s2\rangle$, $\langle t3,s3\rangle$, $\langle t3,s4\rangle$, $\langle t3,s5\rangle$, $\langle t3,s6\rangle$, $\langle t4,s2\rangle$, $\langle t4,s3\rangle$, $\langle t4,s4\rangle$ 按照类型 XiaShang, XiZhou, ChunQiuZhanGuo, QinHan and Cymbals, Bell, Bo, Drum, Chun, Sheng 进行的分类。

表 1　实例连接的分类关系

| Instance connection | XiaShang 夏商 | XiZhou 西周 | ChunQiuZhanGuo 春秋战国 | QinHan 秦汉 | Cymbals 铙 | Bell 钟 | Bo 鎛 | Drum 鼓 | Chun 錞 | Sheng 笙 |
|---|---|---|---|---|---|---|---|---|---|---|
| ⟨t1,s1⟩ | 1 | | | | 1 | | | | | |
| ⟨t1,s3⟩ | 1 | | | | | | 1 | | | |
| ⟨t1,s4⟩ | 1 | | | | | | | 1 | | |
| ⟨t2,s1⟩ | | 1 | | | 1 | | | | | |
| ⟨t2,s2⟩ | | 1 | | | | 1 | | | | |
| ⟨t2,s3⟩ | | 1 | | | | | 1 | | | |
| ⟨t3,s1⟩ | | | 1 | | 1 | | | | | |
| ⟨t3,s2⟩ | | | 1 | | | 1 | | | | |
| ⟨t3,s3⟩ | | | 1 | | | | 1 | | | |
| ⟨t3,s4⟩ | | | 1 | | | | | 1 | | |
| ⟨t3,s5⟩ | | | 1 | | | | | | 1 | |
| ⟨t3,s6⟩ | | | 1 | | | | | | | 1 |
| ⟨t4,s2⟩ | | | | 1 | | 1 | | | | |
| ⟨t4,s3⟩ | | | | 1 | | | 1 | | | |
| ⟨t4,s4⟩ | | | | 1 | | | | 1 | | |

上表蕴含的约束罗列如下：

XiaShang $\vdash_C$ Cymbals, Bo, Drum

XiZhou $\vdash_C$ Cymbals, Bell, Bo

ChunQiuZhanGuo $\vdash_C$ Cymbals, Bell, Bo, Drum, Chun, Sheng

(下转第 273 页)

获得授权句柄 authHandle-$n$ 和临时共享秘密 $S_{temp-n}$。

(2)攻击者对于需要获得实体 entity-$m$ 的授权数据才能执行的命令,输入 authHandle-$n$ 作为授权句柄,并用临时共享秘密 $S_{temp-n}$ 计算输入参数授权消息认证码 inAuth = HMAC($S_{temp-n}$, inParamDigest || inAuthSetupParams)。

上述攻击中,TPM 根据 authHandle-$n$ 找到的临时共享秘密 $S_{temp-n}$ 计算出的 HM 和攻击者所计算出的 inAuth 是一样的,因而能够通过 TPM 的验证使 TPM 认为该命令得到授权,从而执行攻击者本无权执行的命令。由于 TPM 中的大部分命令都可以通过 OSAP 进行授权执行,这个漏洞的安全隐患是很大的。

3.2.2 OSAP 协议的改进

基于以上分析,对 OSAP 协议的实施可以做如下修改:

(1)执行命令 TPM_OSAP( )的过程中,同时保存临时共享秘密 $S_{temp}$ 与共享秘密 $S_{UT}$。

(2)执行 OSAP 协议授权验证时,将与命令参数直接关联的实体授权数据和 TPM 执行命令 TPM_OSAP()时保存的共享秘密 $S_{UT}$ 相比较。如果相等,再用 $S_{temp}$ 进行 HMAC 验证。

这样就把共享秘密 $S_{UT}$ 和命令输入参数直接关联起来了,使得该替换攻击无从下手。

**结束语** 对象无关授权协议(OIAP)、特定对象授权协议(OSAP)是 TPM 在可信计算平台上运行的基本协议,确保这些协议的安全运行极为重要。本文对这两个授权协议进行了逻辑描述并对其安全性进行了分析,针对协议的安全隐患提出了相应的改进方法,对于研发 TPM 芯片时的工程实现提供了参考。

## 参 考 文 献


1 谭兴烈. 可信计算平台中的关键部件 TPM. 信息安全与通信保密,2005-02

2 Graeme Proudler (Hewlett Packard). Trusted Computing, TCG Technical Committee,2004(报告)

3 科学技术信息研究所. 中国 TCG 标准进入倒计时[EB/OL]. 2005-11-25

4 TCG Glossary. http://www trustedcomputinggroup. org, 2004 (6):4~7

5 Trusted computing work group. TPM Main Specification. http://www. trustedcomputinggroup. org. 2005

6 范红. 安全协议形式化分析理论与方法:[博士论文]. 解放军信息工程大学,2003-02

7 可信终端成就安全体系. http://www. ccw. com. cn/applic/tech/htm2004/20040613_1340T. asp

8 陈军. 可信平台模块安全性分析与应用:[博士论文]. 中科院,2006-03

9 叶定松. 计算机可信改造与实践. 见:中国信息安全发展趋势与战略高层研讨会文集,2005,6:51~53


(上接第 263 页)

QinHan $\vdash_C$ Bell, Bo, Drum

……

其中 XiaShang $\vdash_C$ Cymbals, Bo, Drum 可以读做两个数据库中夏商时期的乐器有铙、镈和鼓。

**结论** 信息流理论是一门信息科学理论,它是关于信息如何在分布式系统中流动的数学理论模型,为我们研究分布式系统的语义互操作问题提供了理论基础。而分布式系统的分析形成具有相对性,且既可以是具体的,也可以是抽象的,所以它的应用将非常广泛。

## 参 考 文 献


1 Devlin K. Logic and Information. Cambridge,1991

2 Devlin K. Introduction to Channel Theory. ESSLLI 2001, Helsinki, Finland,2001

3 Dretske. Knowledge and the Flow of Information. Oxford. Basil Blackwell,1981

4 Floridi L. What is the Philosophy of Information? Mataphilosophy, 2002,33(1-2): 123~145

5 Barwise J, Seligman J. Information Flow: the Logic of Distributed Systems. Cambridge: Cambridge University Press,1997

6 Checkland P. Systems Thinking, Systems Practice. Chichester: John Wiley & Sons,1981

7 Kent R E. The Information Flow Framework. Starter document for IEEE P1600. 1, the IEEE Standard Upper Ontology Working Group, 2001. http://suo. ieee. org/IFF/

8 Kent R E. The IFF Approach to Semantic Integration. In: the Boeing Mini-Workshop on Semantic Integration, November 2002

9 Kalfoglou Y, Schorlemmer. Using Information Flow Theory to Enable Semantic Interoperability. In:Proceedings of the 6th Catalan Conference on Artificial Intelligence (CCIA '03), Palma de Mallorca, Spain, October 2003

10 Kalfoglou Y, Schorlemmer M. IFMap: an ontology mapping method based on information flow theory. Journal on Data Semantics,2003, 1(1):98~127

11 Liu Hongzhe,Bao Hong,Feng Junkang. IF based Semantic Interoperability for Distributed Digital Museums. Computing and Information System Journal,2006,10

12 Liu Hong-zhe, Bao Hong, Wu Jing, et al. An Information Flow Based Approach to Semantic Integration of Distributed Digital Museums. ICMLC06,Dalian,China,2006