Two Tier Ontology Alignment

First A. Hongzhe Liu, Second B. Hong Bao, and Third C. Junkang Feng

Abstract**—**It is desirable to the distribute data and knowledge of a huge amount of cultural heritage and to make them available readily to people, but the distributed, heterogeneous and autonomous nature of the databases of digital museums gives rise to the challenge of achieving the best retrieval results in cross-system searching. To make this difficult task tractable, we analyze semantic heterogeneities among these data sources and describe an two tier (concept and instance) approach to accomplish ontology alignment by using the information flow theory.

Keywords**—**Digital museum, information flow, ontology alignment, precision, recall.

I. INTRODUCTION

IN the open environment of the Internet and the Web, information resources are heterogeneous and are indexed with different vocabularies and organized according to different schemes. How to achieve the best retrieval results in cross-domain searching has presented a particular challenge to the information profession. In information retrieval, users typically are neither, nor should they need to be, aware of the behind-the-scenes mechanisms for matching their query terms to the vocabularies employed by various systems. The ideal approach would be to provide a “one-stop” seamless searching instead of requiring the user to search individual databases or collections separately. To enable such an approach, it is important to render the different knowledge organization systems interoperable.

Chinese civilization has been around for thousands of years, and as a result a huge amount of cultural heritage and antiques are scattered all over the vast territory of China. All kinds of digital museums have been developed for them. Each of these museums maintains large digital archives of their collections. They represent an extremely valuable cultural heritage resource, and yet access and exploitation of the data is constrained due to the distributed and heterogeneous nature of the resource. Therefore it is highly desirable to find an avenue to integrating these distributed and heterogeneous systems.

Manuscript received June 9, 2007; Revised received November 10, 2007. This work was supported in part by *the Beijing Computer Science and Technology* emphasis construction subject .

Hongzhe Liu is with the Institute of Information Technology,Beijing Union University, BeiSiHuan East Road, Beijing,CHINA,100101,xxtliuhongzhe@buu.com.cn.

Hong Bao is with the information college, Beijing Union University, BeiSiHuan East Road, Beijing,CHINA,100101,baohong@buu.com.cn.

Junkang Feng is with the School of Computing University of the West of Scotland,UK, junkang.feng@uws.ac.uk

The remainder of this paper is organized as follows: Section 2 describes a motivating example in digital museum domain. Section 3 defines the ontology mapping. Section 4 lists the related works. Section 5 proposes our solution. Finally section 6 summarizes our contribution of the work and gives a conclusion.

II. MOTIVATING EXAMPLE

Given two digital museums 1 and 2 with local ontology alignment as the follows:

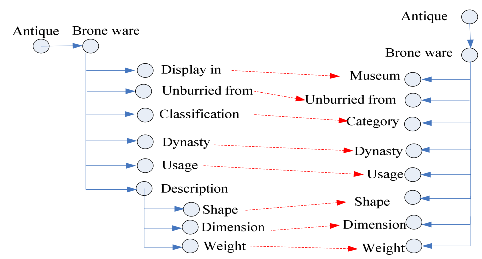

Fig.1 conceptual alignment between two local ontologies

Table 1 Instances of digital museum 1

Bronze Ware title	**Dynasty (*instance values*)**
NiuShouWenLiNiao Bo牛首纹立鸟镈	XiaShang夏商
JuanYunWen Nao卷云纹铙	XiaShang夏商
ShenRenWenShuangNiao Gu神人纹双鸟鼓	XiaShang夏商
JiaYiBingDing Bo鎛甲乙丙丁	XiZhou西周
HuWenZheng Nao虎纹钲（铙）	XiZhou西周
YingHouJianZhong应侯见钟	XiZhou西周
YunWen Nao云纹铙	ChunQiuZhanGuo春秋战国
ZengHouYiBian Zhong曾侯乙编钟	ChunQiuZhanGuo春秋战国
HuNiuYu Chun虎钮錞于	ChunQiuZhanGuo春秋战国
ZhanGuo Gu战国鼓	ChunQiuZhanGuo春秋战国
QuGuanHuLu Sheng曲管葫芦笙	ChunQiuZhanGuo春秋战国
QinGong Bo秦公鎛	ChunQiuZhanGuo春秋战国
Gong Bo公鎛	QinHan秦汉
Gong Zhong公钟	QinHan秦汉
TieJinTong Gu贴金铜鼓	QinHan秦汉
…	…

Table 2 Instances of digital museum 2

Bronze Ware title	**Dynasty (*instance values*)**
JuanYunWen Nao卷云纹铙	Shang商
BianXingShouMianWen Nao变形兽面纹铙	ChunQiu春秋
JinHouMu Zhong晋侯木钟	XiZhou西周
WangSunHaoYong Zhong王孙诰甬钟	Late ChunQiu春秋晚期
Gong Zhong公钟	QinHan秦汉
Xing Zhong兴钟	XiZhou西周
NiuShouWenLiNiao Bo牛首纹立鸟镈	Shang商
PanSheWenBian Bo蟠蛇纹编鎛	Late ChunQiu春秋晚期
Gong Bo公鎛	Qin秦
JiaYiBingDing Bo鎛甲乙丙丁	XiZhou西周

ShenRenWenShuangNiao Gu神人纹双鸟鼓	Late Shang商晚期
WuZhuXianWen Gu五铢线纹鼓	Late XiHan西汉晚期
ZhanGuo Gu战国鼓	ZhanGuo战国
HuNiuYu Chun虎钮錞于	ZhanGuo战国
QuGuanHuLu Sheng曲管葫芦笙	ZhanGuo战国
…	…

Two standards for classifying Bronze Ware are as follows:

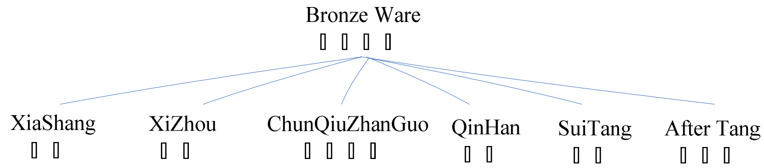

Fig. 2 bronze Ware classifications 1 according to creation time

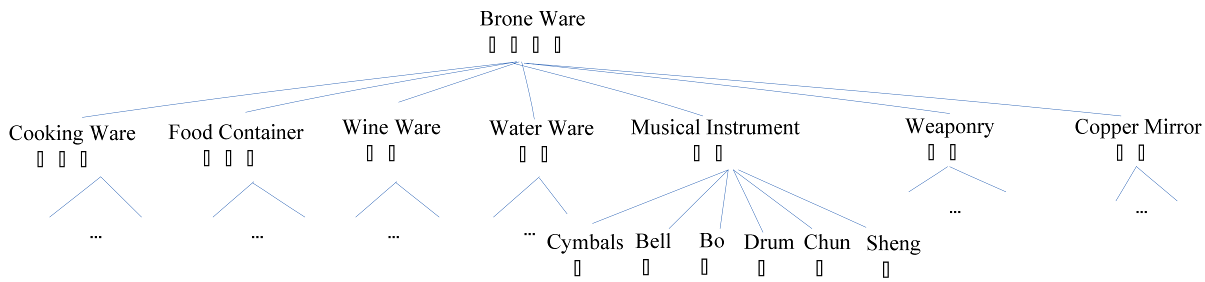

Fig. 3 Bronze Ware classifications 2 according to usage and shape

The category in table 1 is based on the creation time of the Bronze Ware , that is, the dynasty when it was created. The category in table 2 is based on the usages of the Bronze Ware on the first classification level. For example, if it is usually used for cooking food, a piece of Bronze Ware belongs to Cooking Ware. The Musical Instruments are classified further according to their shapes on the second level.

Two local ontology maps at conceptual level in figure 1, and they still cannot communicate with each other due to they have different *instance value*s (shown in tables 1 and 2 and fig. 2 and 3).

III. Problem Definitions

In order define our problem, we give the following definition:

Definition 1 (Ontology) An *ontology* is a pair $\boldsymbol{O} = <\boldsymbol{C}, \boldsymbol{R}>$ where $\boldsymbol{C}$ is a finite set of *concept* connected by R which is a finite set of *relations.*

Definition 2 (conceptual alignment) A *conceptual alignment* $\boldsymbol{B} = <\boldsymbol{C}, \boldsymbol{B}>$, consists of a collection of *concepts* $\boldsymbol{C}=\{\boldsymbol{C}_i\}\ (i\in I)$ from *ontology* $\boldsymbol{O} = \{\boldsymbol{O}_i\}\ (i\in I)$ and a collection of *alignment bridge* $\boldsymbol{B} = \{\boldsymbol{B}_{ij}\}\ i\neq j\in I$ between them.

Definition 3 (ontology alignment) *Ontology alignment* is the process that not only include *the conceptual alignment bridge*, also the ontology *instance*s can be transformed between each other according to their *conceptual alignment.*

As shown with the motivating example in Section 2 when we looked at mappings between local ontologies, normally we can identify approximate *conceptual alignment* between two *ontology* based on their literal meaning. But the source ontology *instance*s cannot transformed into the target ontology entities directly due to their take different *instance values*. *Information flow* theory proposes a principle where the two tiers involved in a *classification*, namely the *types* and the *particulars* both contribute to the flow of information. Drawing on this fundamental idea and that semantic inter-operability is essentially a problem of *information flow*, we propose an two tier ontology alignment approach to distributed digital museums.

IV. Related Works

A project similar to ours is the SCULPTEUR[1] project, which is an ontological-based solution for navigating, searching and retrieving digital cultural heritage information from multiple distributed digital museums. The nature of the work would seem an ontology based semantic integration for multiple heterogeneous sources such as the work of RDFT: mapping meta-ontology for integration task[2], MAFRA: mapping framework for distributed ontologies[3], OntoMerge: Ontology Translation by merging and reasoning[4], Observer: interoperation across preexisting ontologies[5], OIS: framework for ontology integration system[6], somewhere: querying distributed ontologies[7].

Marco Schorlemmer and Yannis Kalfoglou propose an information-theoretic foundation for progressive ontology alignment in [8]. Ontology based integration can overcome semantic heterogeneities, but there is not a universal ontology mapping method for all solutions especially for some specific domain. For example, the classification heterogeneity, which is a commonplace in Chinese antique classifications, would not seem solvable by currently available ontology based methods. In order to find the internal correspondences between different classifications, we propose an *information flow* based solution, which classifies as *types* the hierarchy of the classifications and as *tokens* the *instance*s that satisfy the corresponding *types*, and an agreed *classification* with an indexed family of *infomorphisms* act as *core connections* of them.

V. Using Information Flow For Ontology Alignment

A. Information Flow Theory

Information flow is a rigorous, mathematic theory of knowledge distribution proposed by Barwise and Seligman in 1997[9]. It helps solve problems of how information about one or more components of a distributed system carries information about other components.

Classification. A *classification* is a structure $\boldsymbol{A} = < tok(A), typ(A), \models A >$, where *tok(A)* is a set of objects to be classified, called the *tokens* of ***A***, *typ(A)* is a set of objects used to classify the *tokens*, called the *types* of ***A***, and $\models_A$ is a binary relation between *tok(A)* and *typ(A)* that determines which *tokens* are classified by which *types*. If $a \models_A \alpha$ then we say that *a* is of type α in *classification* ***A***.

Theory. Given a *classification* ***A***, a *sequent* is a pair (Γ, Δ) of sets of *types* of ***A***.

A *token a* of ***A*** is said to satisfy the *sequent* (Γ, Δ) if,

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

We say that Γ *entails* Δ in ***A***, written $\Gamma \vdash_A \Delta$, if every *token* of ***A*** satisfies (Γ, Δ).

If $\Gamma \vdash_A \Delta$, then the pair (Γ, Δ) is said to be a *constraint* supported by the *classification* ***A***.

The set of all *constraints* supported by ***A*** is called the *complete theory* of ***A***, denoted by *Th(A)*. The complete *theory* of *A* represents all the regularities supported by the system being modeled by ***A***.

- **Entailment:** a *constraint* of the form $\alpha \vdash \beta$ represent the claim that α *entails* β.
- **Necessity:** a *constraint* of the form $\vdash \alpha$ represent the claim that the type α is necessarily the case, without any preconditions.
- **Exhaustive:** a *constraint* of the form $\vdash \alpha, \beta$ represent the claim that every *token* is of one of the two *types* α and β, again without any preconditions.
- **Incompatible types:** a *constraint* of the form $\alpha, \beta \vdash$ represents the claim that no *token* is of both *types* α and β.
- **Incoherent types:** a *constraint* of the form $\alpha \vdash$ represents the claim that no *token* is of type α.

Infomorphism. Let $\boldsymbol{A} = < tok(A), typ(A), \models A >$, and $\boldsymbol{C} = < tok(C), typ(C), \models C >$, be two *classifications*. An *infomorphism* between ***A*** and ***C*** is a contravariant pair of function $f = (f^{\wedge}, f^{\vee})$ that satisfies the following Fundamental Property of *Infomorphism*:

$$f^{\vee}(c) \models_A \alpha \quad iff \quad c \models_c f^{\wedge}(\alpha)$$

For all *tokens c* of *C* and all *types α* of *A*. We refer to $f^{\wedge}$ as "f-up" and $f^{\vee}$ as "f-down". We take account of the fact that the functions $f^{\wedge}$ and $f^{\vee}$ act in opposite directions by writing $f : A \rightleftarrows C$.

Local logics A *local logic* $\boldsymbol{L} = <\boldsymbol{A}, \vdash_L, \boldsymbol{N_L}>$ consists of a *classification* ***A***, a set $\vdash_L$ of *sequents* (satisfying certain structural rules) involving the *types* of ***A***, called the *constraints* of *L*, and a subset $\boldsymbol{N_L} \subseteq \boldsymbol{A}$, called the *normal tokens* of *L*, which satisfy all the *constraints* of $\vdash_L$. A *local logic L* is **sound** if every *token* is *normal*; it is **complete** if every *sequent* that holds of all *normal tokens* is in the consequence relation $\vdash_L$. A *logic* is **natural** if it is generated by some *classification*, and a *natural logic* is *sound* and *complete*. Using *infomorphisms*, we can move *local logics* around from one *classification* to another even though this normally preserve neither *soundness* nor *completeness*.

If Γ is a set of *types* of ***A***, we denote by Γ^f the set of translations of *types* in Γ. If Γ is a set of *types* of ***B***, we denote by Γ^{-f} the set of *types* of ***A*** whose translations are in Γ.The following two inference rules allow us to pass from one *classification* to another:

$$\textbf{f-Intro:} \quad \frac{\Gamma^{-f} \vdash_A \Delta^{-f}}{\Gamma \vdash_B \Delta}$$

$$\textbf{f-Elim:} \quad \frac{\Gamma^{f} \vdash_B \Delta^{f}}{\Gamma \vdash_A \Delta}$$

The first rule allows us to go from a *sequent* of ***A*** to a *sequent* of ***B***; the second rule allows us to go the other way round. It can be proven (which is omitted in this paper and interested user is referred to Barwise and Seligman 1997) that *f*-**Intro** preserves *soundness*, whereas *f*-**Elim** preserves *completeness*.

Given an *infomorphism* f: $\boldsymbol{A} \rightleftarrows \boldsymbol{B}$ and a *logic L* on one of these *classifications*, we obtain a *natural logic* on the other. If *L* is a *logic* on ***A***, then $f[L]$ denotes the *logic* on ***B*** obtained from *L* by *f*-**Intro**. If *L* is a *logic* on ***B***, then $f^{-1}[L]$ denotes the *logic* on ***A*** obtained from *L* by *f*-**Elim**.For any binary *channel* ***C*** as Fig.4, we define the *local logic* $\mathrm{Log}_C(\boldsymbol{D})$ on ***D*** induced by that *channel* as $\mathrm{Log}_C(\boldsymbol{D}) = d^{-1}[p[\mathrm{Log}(\boldsymbol{P})]]$

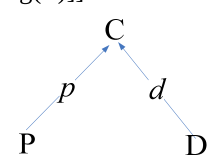

Fig. 4 Binary logic channel C

As we have observed, if Log (***P***) is *sound* and *complete*, and p、d is *token surjective*, then $\mathrm{Log}_C(\boldsymbol{D})$ is *sound* and *complete*.

B. Proposed Solution

Followed by above, *C* connect *P* and *D* by means of two pairs of contra-variant functions, it captures an existing duality between *concept*s and *instance*s: Each pair consists of a map of concepts on the so called *type* level and map of *instance*s on the so called *token* level, and pointing in the opposite direction. We model the two Bronze Ware categories shown in figure 2 and figure 3 as ***P(C) and D(C)***, and they are connected by an agreed understanding ***C(C)***, the *type*, *token* and generated *theory* details of the *two local logics* are illustrated bellow. The diagram is a general model of the coordinated channel between two local ontology classifications, and it faithfully captures the semantic alignment between them, according to the Barwise-Seligman theory of *information flow*. With new tokens are shared, new types participate in the meaning coordination, and yielding the newly coordinated channel. The distributed IF logic of the natural logic determined by the core of each new channel captures the semantic integration achieved so far. Because usually the token information is not as complete as needed or its computational cost is too expensive, in practice complete semantic integration will seldom be achieved, the ontology coordination process will usually yield only a partial semantic integration involving a fraction of communicated types and shared tokens.

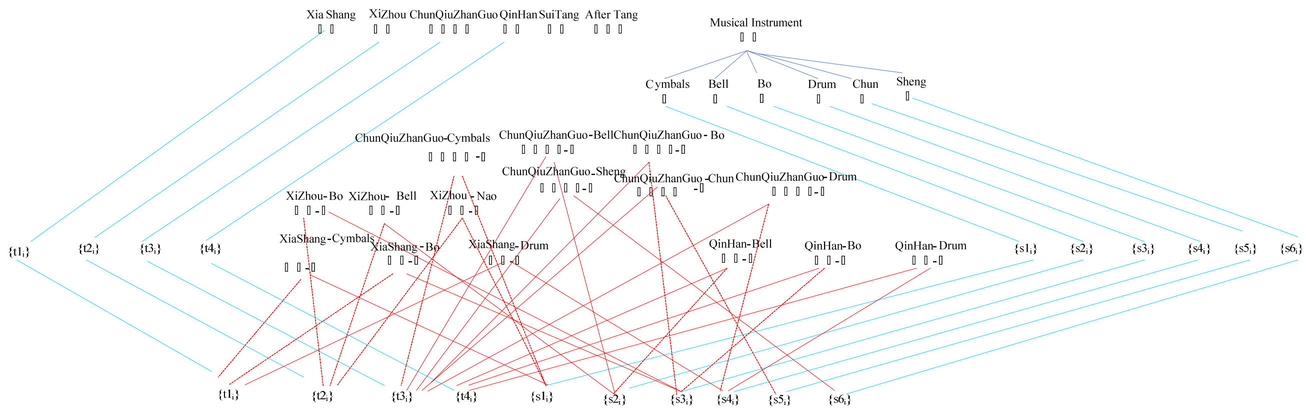

Fig.5 Bronze ware classification alignment

P(C):Type: *XiaShang, XiZhou, ChunQiuZhanGuo, QinHan* **Token:** $\{t1_i\},\{t2_i\},\{t3_i\},\{t4_i\}$

Theory : $\vdash_{P(C)}$ *XiaShang, XiZhou, ChunQiuZhanGuo, QinHan ; XiaShang, XiZhou, ChunQiuZhanGuo, QinHan* $\vdash_{P(C)}$

D(C):Type: *Cymbals, Bell, Bo, Drum, Chun, Sheng* **Token:** $\{s1_i\}$, $\{s2_i\}$, $\{s3_i\}$, $\{s4_i\}$, $\{s5_i\}$, $\{s6_i\}$ **Theory:** $\vdash_{D(C)}$ *Cymbals, Bell, Bo, Drum, Chun, Sheng;Cymbals, Bell, Bo, Drum, Chun, Sheng* $\vdash_{D(C)}$

C(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:** $<\{t1_i\},\{s1_i\}>$, $<\{t1_i\},\{s3_i\}>$, $<\{t1_i\},\{s4_i\}>$, $<\{t2_i\},\{s1_i\}>$, $<\{t2_i\},\{s2_i\}>$, $<\{t2_i\},\{s3_i\}>$, $<\{t3_i\},\{s1_i\}>$, $<\{t3_i\},\{s2_i\}>$, $<\{t3_i\},\{s3_i\}>$, $<\{t3_i\},\{s4_i\}>$, $<\{t3_i\},\{s5_i\}>$, $<\{t3_i\},\{s6_i\}>$, $<\{t4_i\},\{s2_i\}>$, $<\{t4_i\},\{s3_i\}>$, $<\{t4_i\},\{s4_i\}>$ **Theory:** $\vdash_{C(C)}$*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;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* $\vdash_{C(C)}$

Fig. 5 shows how both categories base their *concept*s upon their agreed understanding. For example, the agreed understanding is materialized by two maps that form the alignment. It requires the *classification* of particular *instance*s of both categories according with the agreed understanding, since it is this agreed way of *classification* which will determine how the both categories are going to be related to each other. The following table 3 and 4 shows connections classified into the *types* which determines a *theory* of how these *concept*s are related.

Table 3 Connections classified according to *types* from one point of view

Instance connection	*Xia Shang*	*Xi Zhou*	*ChunQiu ZhanGuo*	*Qin Han*	*Cymbals*	*Bell*	*Bo*	*Drum*	*Chun*	*Sheng*
$<\{t1_i\},\{s1_i\}>$	1				1					
$<\{t1_i\},\{s3_i\}>$	1						1			
$<\{t1_i\},\{s4_i\}>$	1							1		
$<\{t2_i\},\{s1_i\}>$		1			1					
$<\{t2_i\},\{s2_i\}>$		1				1				
$<\{t2_i\},\{s3_i\}>$		1					1			
$<\{t3_i\},\{s1_i\}>$			1		1					
$<\{t3_i\},\{s2_i\}>$			1			1				
$<\{t3_i\},\{s3_i\}>$			1				1			
$<\{t3_i\},\{s4_i\}>$			1					1		
$<\{t3_i\},\{s5_i\}>$			1						1	
$<\{t3_i\},\{s6_i\}>$			1							1
$<\{t4_i\},\{s2_i\}>$				1		1				
$<\{t4_i\},\{s3_i\}>$				1			1			
$<\{t4_i\},\{s4_i\}>$				1				1		

We can conclude with the following *constraints* according to above table, which is the *theory* about relations between the two antique categories:

XiaShang $\vdash_{C(C)}$ *Cymbals, Bo, Drum;XiZhou* $\vdash_{C(C)}$ *Cymbals, Bell, Bo;ChunQiuZhanGuo* $\vdash_{C(C)}$ *Cymbals, Bell, Bo, Drum, Chun, Sheng;QinHan* $\vdash_{C(C)}$ *Bell, Bo, Drum*

Table 4 Connections classified according to *types* from another point of view

Instance connection	*Cymbals*	*Bell*	*Bo*	*Drum*	*Chun*	*Sheng*	*Xia Shang*	*Xi Zhou*	*ChunQiu ZhanGuo*	*Qin Han*
$<\{s1_i\},\{t1_i\}>$	1						1			
$<\{s1_i\},\{t2_i\}>$	1							1		
$<\{s1_i\},\{t3_i\}>$	1								1	
$<\{s2_i\},\{t2_i\}>$		1						1		
$<\{s2_i\},\{t3_i\}>$		1							1	
$<\{s2_i\},\{t4_i\}>$		1								1
$<\{s3_i\},\{t1_i\}>$			1				1			
$<\{s3_i\},\{t2_i\}>$			1					1		
$<\{s3_i\},\{t3_i\}>$			1						1	
$<\{s3_i\},\{t4_i\}>$			1							1
$<\{s4_i\},\{t1_i\}>$				1			1			
$<\{s4_i\},\{t3_i\}>$				1					1	
$<\{s4_i\},\{t4_i\}>$				1						1
$<\{s5_i\},\{t3_i\}>$					1				1	
$<\{s6_i\},\{t3_i\}>$						1			1	

*Constraint*s concluded:

Cymbals $\vdash_{C(C)}$ *XiaShang, XiZhou, ChunQiuZhanGuo;Bell* $\vdash_{C(C)}$ *XiZhou, ChunQiuZhanGuo, QinHan;Bo* $\vdash_{C(C)}$*XiaShang,* *XiZhou, ChunQiuZhanGuo, QinHan;Drum* $\vdash_{C(C)}$ *XiaShang, ChunQiuZhanGuo, QinHan;Chun* $\vdash_{C(C)}$ *ChunQiuZhanGuo;Sheng* $\vdash_{C(C)}$ *ChunQiuZhanGuo*

Again, we model two 'dynasty' *concepts*' *instance values* and their *instances* classified according to the values as two *classifications*, and their generated *theories* as ***P(D)*** and ***D(D)*** separately, with ***C(C)*** as their *core connection*. The alignment between two 'dynasty' *concepts* is shown in Fig.6:

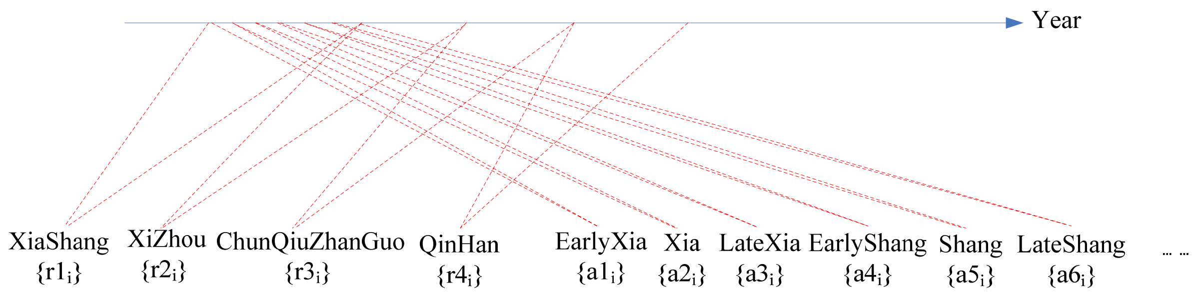

Fig.6 Two 'dynasty' *concepts* alignment

P(D):Type: *XiaShang XiZhou ChunQiuZhanGuo QinHan***Token:**$\{r1_i\}, \{r2_i\}, \{r3_i\}, \{r4_i\}$ **Theory:** $\vdash_{P(D)}$*XiaShang XiZhou ChunQiuZhanGuo QinHan;XiaShang XiZhou ChunQiuZhanGuo QinHan* $\vdash_{P(D)}$

D(D) :Type:*EarlyXia,Xia,LateXia,EarlyShang,Shang,LateShang…***Token:** $\{a1_i\}, \{a2_i\}, \{a3_i\}, \{a4_i\}, \{a5_i\}, \{a6_i\}$ **Theory:** $\vdash_{D(D)}$*EarlyXia,Xia,LateXia,EarlyShang,Shang,LateShang …;EarlyXia,Xia,LateXia,EarlyShang,Shang,LateShang…* $\vdash_{D(D)}$

C(D) :Type:*XiaShang-EarlyXia,XiaShang-Xia,XiaShang-LateXia,XiaShang-EarlyShang,XiaShang-Shang,XiaShang-LateShang...***Token:**$<\{r1_i\}, a1_i\}><\{r1_i\}, \{a2_i\}><\{r1_i\}, \{a3_i\}><\{r1_i\}, \{a4_i\}><\{r1_i\},\{a5_i\}><\{r1_i\},\{a6_i\}>…$

Theory:*EarlyXia* $\vdash_{C(D)}$*XiaShang,Xia* $\vdash_{C(D)}$*XiaShang,LateXia* $\vdash_{C(D)}$*XiaShang,EarlyShang* $\vdash_{C(D)}$*XiaShang,Shang* $\vdash_{C(D)}$*XiaShang,LateShang* $\vdash_{C(D)}$*XiaShan;…;*

XiaShang $\vdash_{C(D)}$*EarlyXia,Xia,LateXia , Early Shang, Shang, LateShang;…*

C. Ontology alignment process

Mutual alignment process

Two ontologies are randomly chosen from the local ontology population. One is called *O* and the other is called *O'*. *O* has a *concept* set $\{C_i\}(i\in [1,m]$, *m* is *concept* number of *ontology O*) and *O'* has a *concept* set $\{C'_j\}(j\in [1,n]$, *n* is concept number of *ontology O'*), C_i and C'_j are two equivalent *concepts* from *O* and *O'*, C_i has an *instance value* set $\{c_k\}(k\in [1,M]$, *M* is number of *instance value* of C_i) and C'_j has an *instance value* set $\{c'_t\}(t\in [1,N]$, *N* is number of *instance value* of C'_j). Every *instance value* c_k may populate with a group of *instance* set $\{o_{ks}\}(s\in [1,t]$, *t* is number of *instance* of c_k) and c'_t with $\{o'_{tr}\}(r\in [1,T]$, *T* is number of *instance* of c'_t). The ontology alignment process is executed repeatedly in a (finite or infinite) sequence of rounds. One round for each *instance value* of *concept* C. On each round, the following events happen:

1.C_i chooses one *instance value* c_k from its *instance values* set $\{c_k\}$, c_k then chooses an *instance* o_{ks} from *instance* set $\{o_{ks}\}$ which has an *instance value* c_k , denoted as $o_{ks}\in c_k$.

2.C'_j receives the *instance* o_{ks}, and find which *instance value*(s) c'_t from $\{c'_t\}$ it belongs to, if it is failed to find any *instance value*, then end this round;

3.For every locating *instance value* $c'_t \in C'_j$ such that $o_{ks} \in c'_t$: update it in $C_i\&C'_j$ *association matrix*, then *concept* C'_j select every object o'_{kr} from its *instance value* set $\{ o'_{tr} \}$ which is different from o_{ks}, and sends them back to C_i to indicate that o_{ks} and o'_{tr} all belong to the same *instance value* in c'_t.

4.C_i receive o'_{tr} and judge it whether match any its *instance values* of it, if not, do nothing. if yes, update this in $C_i\&C'_j$ *association matrix*;

Association matrix

In our model, we assume that equivalent *concepts* keeps an *association matrix*, which is used to capture relationships among objects. After each round of the alignment, the *association matrix* might be updated to reflect the new relationships among objects. Following are the concept *association matrix S[m][n]* of equivalent *concept* C_i from *O* and C'_j from *O'*, *concept* C_i has *instance values* $c_1,c_2,c_3,……c_n$ whereas *concept* C'_j has *instance value* $c'_1,c'_2,……c'_m$:

s_{ij} is value of the ssociation matrix between two concept instance values:

s_{ij} = { *o* |collection of instances that belongs to instance value c_i and c'_j }

Association strength

Denote $AS(c_i,c'_j)$ as *association strength* between two *instance value* c_i and c'_j.:

$AS(c_i,c'_j)$ = | $\boldsymbol{s_{ij}}$ |, number of *instances* of s_{ij}

Given local *Ontology O* and *O'* have a set of *concepts* $\{C_i\}$ and $\{C'_j\}$ separately, the *association strength* of *concept* C_i and C'_j denote as $AS(C_i,C'_j)$:

$$AS(C_i,C'_j)=\sum AS(c_i,c'_j) = \sum_{i=1}^{m}\sum_{j=1}^{n} |\, \boldsymbol{s_{ij}} \,|$$

The association strength of *O* and *O'* denote as $AS(O,O')$:

$AS(O,O')=\sum_{k=1}^{s}\omega_k \cdot AS(C_i,C'_j)$, s is number of equivalent *concept*s exist in *ontology* O and O'. ω_k is the *weight* of the *association strength* of *concept* C_i and C'_j ,it can be defined by the need of application.

We can set *threshold* value for the *association strength* $AS(c_i,c'_j)$, $AS(C_i,C'_j)$, $AS(O,O')$, if their value are greater than the *threshold*, then the involved *instance values*, *concepts* or *ontology* can be treated as '*relevant*'.

Two tier ontology alignment

The two tier local ontology alignment is as following fig. 7. It does the ontology alignment between *concepts* and *instance values*.

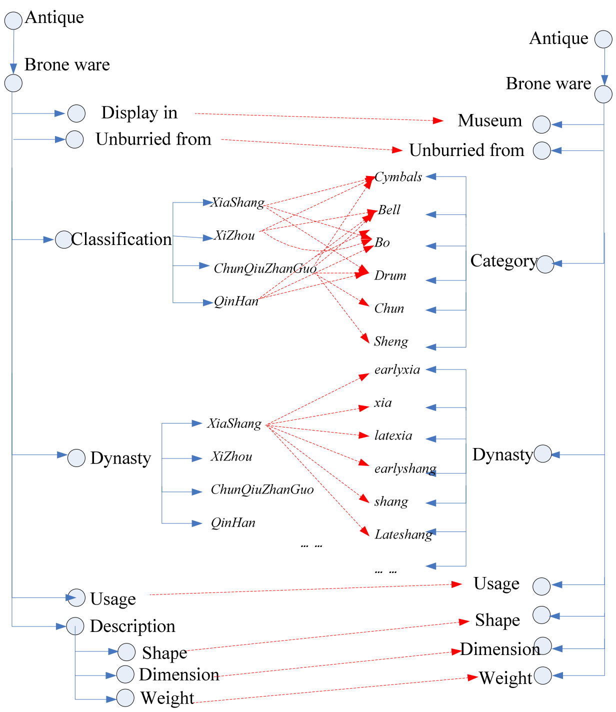

Fig.7 Two tier ontology alignment

D. Peer to peer query application

Peer to peer query architecture

We have constructed a prototype system that involves querying distributed heterogeneous ontologies, the system consists of a collection of peers that maintain local ontologies, including repositories of data instantiating certain *concept*s of local ontologies. Peers are acquainted with neighbors via peer alignment algorithm, that is a set of mappings rules, the mapping rules are formulated according to *association strength* computed of the local *ontology* peers, there is a collection of *association strength values* between peer *ontology concept*s and *concept instance value*s. if the association strength value of one *ontology* peer is higher the threshold value, then a query is passed to its neighbor and relevant *concept*s and *instance value*s are mapped according to *association strengths* between them. Queries are posed to a given peer using one local *ontology* of the peer. The answers that are expected are not only *instance*s of given peer classes but also *instance*s of classes of foreign peers obtained via specified mappings.

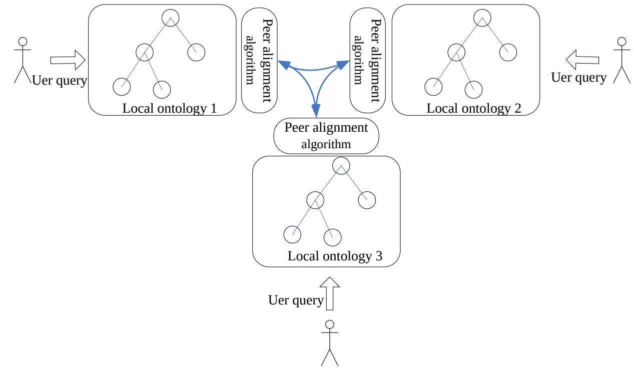

Fig.8 Distributed peer-to-peer query architecture

Precision/recall and the association strength threshold value

Precision and *Recall* have been very widely used in Information Retrieval literature to measure loss of information incurred when the answer to a query issued to the information retrieval system contains some proportion of irrelevant data.

Precision=proportion of the retrieved objects that are relevant

=Probability(Relevant/Retrieved)

$$=\frac{RetrievedSet \cap RelevantSet}{RetrievedSet}$$

Recall=proportion of the relevant objects that are retrieved

= Probability(Retrieved/Relevant) =

$$\frac{RetrievedSet \cap RelevantSet}{RelevantSet}$$

Learn from practice, we know that the higher(lower) the threshold values for the *association strength* $AS(c_i,c'_j)$, $AS(C_i,C'_j)$ and $AS(O,O')$ we set, the higher(lower) of the *Precision*, but the lower(higher) of the *Recall*. We can adjust the *threshold value* according to the application to achieve optimum result.

VI. Conclusion And Contribution

Information flow emphasises that, since information is carried by particular tokens, *information flow* crucially involves both types and tokens. Barwise and Seligman realised the fundamental duality between types and tokens, which is central to all channel theoretic constructions. Thus, although meaning coordination is usually thought of as a process during which *concept*s of separate ontologies are being aligned at the type-level, the logical relationship between *concept*s arises when tokens are being connected by means of an IF channel. Knowing what these connections at the token-level are is therefore fundamental for determining the semantic alignment of ontologies at the type-level.

In this paper, we proposed an two tier ontology alignment algorithm for heterogeneous digital museums and show why conceptual level alignment is insufficient in semantic interoperability. We construct our distributed peer-to-peer query architecture based on that.

The problem of decentralized querying distributed data through distributed ontologies is that the *Precision* and *Recall* is hard to measure and control, experiment proves that our *information flow* based two tier ontology alignment solution can optimize the *Precision* and *Recall* rate as well as make them measurable by adjusting the *threshold value* of *association strength*.

REFERENCE

[1] B. Omelayenko. RDFT: A Mapping Meta-Ontology for Business Integration.In Proceedings of the Workshop on Knowledge Transformation for the Semantic Web, *the 15th European Conference on Artificial Intelligence (KTSW-2002)*, pages 77–84, 2002.

[2] A. Maedche, B. Motik, N. Silva, and R. Volz. MAFRA - a Mapping Framework for Distributed Ontologies. *Proceedings of the Knowledge Engineering and Knowledge Management (EKAW-2002),* volume 2473 of Lecture Notes in Computer Science. Springer, 2002.

[3] [22] D. Dou, D. McDermott, and P. Qi. Ontology Translation on the Semantic Web. *The Journal on Data Semantics,* 3360:35–57, 2004.

[4] E. Mena, A. Illarramendi, V. Kashyap, and A. Sheth. OBSERVER: An Approach for Query Processing in Global Information Systems based on Interoperation across Pre-existing Ontologies. *International Journal on Distributed and Parallel Databases,* 8(2):223–272, 2000.

[5] D. Calvanese, G. De Giacomo, andM. Lenzerini. A Framework for Ontology Integration. *Proceedings of the 2001 International Semantic Web Working Symposium (SWWS-2001)*, pages 303–316, 2001.

[6] F. Goasdou´e and M-C. Rousset. Querying Distributed Data through Distributed Ontologies: a Simple but Scalable Approach. *IEEE Intelligent Systems*, 18(5):60–65, 2003.

[7] Schorlemmer, M. and Kalfoglou, Y. (2005) *Progressive Ontology Alignment for Meaning Coordination: an Information-Theoretic Foundation.* In: 4th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'05), July, Utrecht, Holland.

[8] J. Barwise and J. Seligman, *Information Flow the Logic of Distributed Systems*, Cambridge Tracts in Theoretical Computer Science 44, 1997.

[9] G.Salton, Automatic text processing, Addison_Wesley.