Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 34 (2008) 2721–2731
Expert Systems with Applications
www.elsevier.com/locate/eswa

Taxonomy alignment for interoperability between heterogeneous virtual organizations

Jason J. Jung *

Department of Computer Engineering, Yeungnam University Dae-Dong, Kyungsan, Korea

Abstract

Resources in virtual organizations are classified based on their local taxonomies. However, heterogeneity between these taxonomies is a serious problem for efficient cooperation processes (e.g., knowledge sharing and querying-based interactions). In order to overcome this problem, we propose a novel framework based on aligning the taxonomies of virtual organizations. Thereby, the best mapping between two organization taxonomies has to be discovered to maximize the summation of a set of partial similarities between concepts in the taxonomies. We can consider two levels of alignment processes; (i) intra-alignment in a virtual organization for building an organizational taxonomy and (ii) inter-alignment between organizational taxonomies. Particularly, for intra-alignment, features extracted from resources are exploited to enhance the precision of similarity measurement between concepts. For experimentation, twelve virtual organizations have been built with different local taxonomies. The proposed inter-alignment method has shown about 76% of precision and 68% of recall. Also, feature-based intra-alignment improved those performance, during resource retrieval by query transformation. In addition, we found out that alignment results are dependent on some characteristics of taxonomies (e.g., depth and number of classes).
© 2007 Elsevier Ltd. All rights reserved.

Keywords: Taxonomies; Alignment; Knowledge sharing; Virtual organizations

1. Introduction

Virtual organizations (VOs) have been regarded as one of the richest open information space in various domains, e.g., e-business, e-learning and digital libraries. Especially, a virtual enterprise (VE) is an ad-hoc and automated coalition between businesses that come together to share skills (and knowledge) or core competencies and resources in order to better respond to business opportunities, and whose cooperation is supported by computer networks. The concept of VE has been applied to many forms of cooperative business relations, like outsourcing, supply chains, or temporary consortium (Lopes Cardoso & Oliveira, 2004).

Basically, a VO is composed of (i) a resource repository, which stores massive resources (e.g., electronic documents, multimedia data and so on), and (ii) multiple information systems, which are providing relevant services and functionalities to process and manage the resources (Abrol et al., 2005; Jonkers et al., 2004). The information systems are playing an important role of supporting decision makers. For making business decisions better, users (e.g., decision makers) have to request various information retrieval-related services (e.g., query processing and data warehousing) to the information systems for accessing to the resource repository.

As an important feature, the information systems can employ their own classification systems like directories, catalogue, and yellow pages (Cilia & Buchmann, 2002; Jung, 2007b). Such approaches are mainly based on taxonomy structure, represented as a hierarchical structure between topics (or classes) in common, for organizing a large amount of resources. For instance, some resources about "virtual enterprise" can be annotated with "*Computers* > *Software* > *Enterprise Information Integration*", i.e., classified into the corresponding classes (Jung, 2007a).

* Present address: Department of Computer Engineering, Yeungnam University Dae-Dong, Kyungsan, Korea 712-749. Tel.: +82 32 875 5863.
E-mail address: j2jung@intelligent.pe.kr

0957-4174/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2007.05.015

However, the problem is that formation of the taxonomies are semantically distinct with each other, because the taxonomies are designed by experiences and heuristics of the local experts (or administrators). It means that semantic information extracted from the taxonomies may be heterogeneous with the others. Such heterogeneities are caused by the difference of not only the terminologies (e.g., synonyms and antonym), but also, more importantly, the knowledge structures (e.g., database schema Hull, 1997 and ontologies Jung, 2006). We note two main semantic heterogeneities between taxonomy-based businesses, as follows.

1. Lexical heterogeneity. Even though the classes of taxonomies are semantically equivalent, keywords used for expressing the classes might be different from other VEs. For example, a class for "*Computer Science*" can be represented as "*CS*" as well as "*Informatiks*". Additionally, this sort of heterogeneities is also caused by (i) multi-lingual problem and (ii) synonyms (or antonyms) (Menczer, 2004).
2. Structural heterogeneity. Semantic relationships (e.g., subclass, superclass, and so on) between two concepts in a taxonomy are different from others. There also exist some missing concepts. For the practical reason, Jung mentions class duplications between identical categories and the subordination between dependent categories (Jung, 2005).

Consequently, the information systems are difficult to be integrated, and more importantly, the VOs are impossible to automatically achieve *strategic* cooperations with heterogeneous VOs.

In order to overcome this drawback, we have focused on semantic interoperability between these taxonomy-based virtual enterprises (VEs). A large number of businesses have been inter-related with the others in a same VE or different VEs for performing ad-hoc (or real-time) collaboration. In order to provide efficient interoperability between the enterprises, the heterogeneities between the corresponding taxonomic knowledge structures have to be dealt with. Thereby, we have to consider efficient alignment method to resolve their conflicts. While intra-alignment is a process merging all local taxonomies into an organizational taxonomy, inter-alignment is a process mapping all semantic correspondences between two organizational taxonomies. As shown in Fig. 1, for the interactions between two virtual enterprises VE_A and VE_B, their organization taxonomies T_A and T_B have to be aligned, in advance. Triangles indicate the corresponding local taxonomies. Two alignments for (i) intra- and (ii) inter-organization are shown as solid arrows and a dotted arrow, respectively.

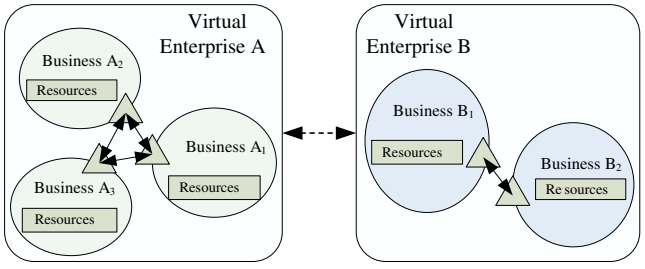

Fig. 1. Interoperability between two virtual enterprises.

Many studies have been proposed to provide interoperability by discovering and integrating local knowledge structures between VOs (Castano, Ferrara, & Montanelli, 2006). They can be briefly noted into three issues;

- Incremental discovery of local knowledge (Jung, 2007b),
- Knowledge matching (including schema and ontology matching) (Shvaiko & Euzenat, 2005), and
- Interoperability via third-party platforms, e.g., service-oriented architecture (SOA) (Guido & Maurizio, 2005).

We propose a novel method to build a VE by mapping heterogeneous taxonomies of businesses, i.e., maximizing the summation of partial similarities between a set of possible pairs of classes. The partial similarity can be calculated by comparing both set of instances in the classes. After both taxonomies are aligned at conceptual level, and the source ontology instances are transformed into the target taxonomy entities according to those semantic relations.

Additionally, we are focusing on supporting local users (e.g., decision makers) through aligning the taxonomies applied to annotate (or classify) the resources on VEs. It means the local users in a certain VE can access to the other VEs which are not familiar with them. Unlike a centralized portal systems (e.g., meta search engines), the local users can be provided a set of topic mapping extracted from direct alignments, so that they can deploy meaningful translation services (e.g., query expansion (Qiu & Frei, 1993) and transformation).

The remainder of this paper is as follows. In the following Section 2, we describe the problem of semantic heterogeneity between VEs. Sections 3 and 4 propose a novel similarity measurement between heterogeneous taxonomies, and alignment-based interoperability applications by using these similarity measurement. In Section 5, experimental results will be shown to evaluate our approach. Section 6 discusses some significant issues and compares our contributions with the previous studies. Finally, Section 7 draws our conclusions of this work.

2. Heterogeneous taxonomies: problem description

For the purpose of managing local resources, VEs can exploit a taxonomy-based classification system to hierarchically organize the resources. The taxonomy is a tree structure composed of a set of classes describing the domain-specific knowledge. We assume that a local organizational taxonomy $\mathcal{T}$ of a VE should be organized as a set of faceted taxonomies and manual alignments by domain experts.

Definition 1 (*Faceted taxonomy*). Let $\mathcal{C}_k$ a set of all concepts of a business B_k, participating in a VE_α. A faceted taxonomy FT_k is defined as a set of subclass assertions between classes in the concept set $\mathcal{C}$. Hence, FT_k is given by

$$\mathrm{FT}_k = \{c_{\mathrm{root}}, \langle c_i, subc, c_j\rangle | c_i, c_j \in \mathcal{C}_\alpha, c_j = \mathrm{subClass}(c_i)\} \tag{1}$$

where c_i means a superclass of c_j. We put c_{root} as root class of FT_k for convenience.

On the top of this structure, a root node is playing a role of simple connector among faceted taxonomies. As drilling down from this root, the classes are more branched and more specified. Each class is containing a subset of instances in its superclass.

Definition 2 (*Instance*). Let $\mathcal{I}_\alpha$ a set of all instances in B_α. A set of instances in class $c_i \in \mathcal{C}_\alpha$ is denoted as $I_i = \{d_1, d_2, \ldots, d_{|I_i|}\} \subseteq \mathcal{I}_\alpha$.

In this paper, we assume that the instances (e.g., textual documents and multimedia data) should be annotated with semantic information from the taxonomies.

Definition 3 (*Organizational taxonomy*). An organizational taxonomy $\mathcal{T}$ in a VE is built by aggregating a set of faceted taxonomies. Thus, supposing that a set of businesses $\{B_1, \ldots, B_{|\alpha|}\}$ be comprised in VE_α, organizational taxonomy $\mathcal{T}_\alpha$ is formulated by

$$\mathcal{T}_\alpha = \bigcup_{B_k \in \mathrm{VE}_\alpha} \mathrm{FT}_k \tag{2}$$

where c_{root} in all FT are equivalently aligned. More importantly, domain experts can manually assert alignments $\mathcal{A}_\alpha = \{\langle c_p, rel^\star, c_q\rangle^\star | c_p \in \mathrm{FT}_p, c_q \in \mathrm{FT}_q\}$. These mappings are expressed with various relations between classes in different faceted taxonomies.

For building each business's taxonomy, they can import the faceted taxonomies, and aggregate them. For such relations *rel*, this paper considers only subclass, superclass, and equivalence. Fig. 2 shows a simple example of business taxonomies. Root classes are simply overlapped (i.e., blue dotted line). Three pairs of classes (i.e., red dotted lines) are explicitly aligned by human experts.[1]

At the moment, the only way to take advantage of the instances in other taxonomies is to get cross through the manual alignments provided by human experts. This kind of alignments however is based on time-wasting tasks. They have to realize and understand the semantic structures of given taxonomies. Such tasks are (i) to scan most of instances in each class (what kinds of instances are included in classes), and (ii) to reflect their own experiences and heuristics (which relations are involved between classes).

[1] For interpretation of color in Fig. 2, the reader is referred to the web version of this article.

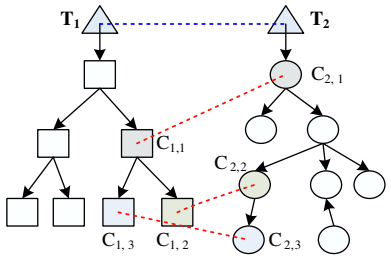

Fig. 2. An example of a business taxonomy.

More seriously, there can exist some missing alignments, which might be significant. For the interoperability between taxonomy-based VEs, at least, the minimum number of class matchings should be asserted. We are considering that the ratio of alignment $\rho_{\mathrm{Align}}(\alpha) = \frac{|\mathcal{A}_\alpha|}{\sum_{k=1,\mathrm{FT}_k\in\mathcal{T}_\alpha}^{k}|\mathrm{FT}_k|}$ should be another important factor for the quality of taxonomy alignment. This ratio factor should be compared with user-specified threshold $\tau_{\mathcal{A}}$. For example, let $\tau_{\mathcal{A}} = 0.3$ in VE_β, and two faceted taxonomies FT_a, FT_b be given to align into $\mathcal{T}_\beta$. If $\rho_{\mathrm{Align}}(\beta) \leqslant \tau_{\mathcal{A}}$, the VE_β is hard to execute not only internal operations but also efficient collaborations with other VEs. We will discuss how to obtain optimal value of this factor in Section 5.

3. Taxonomy alignment

In order to solve these drawbacks, discovery process for significant alignments between taxonomies needs to be automated. A set of given taxonomies have to be matched as finding out the best configuration of alignments between classes. We assume that the best configuration should be maximizing the summation of class similarities. Similarity between two classes is computed by not only class labels but also features extracted from the instances.

3.1. Feature mapping from manual alignments

We exploit the set of alignments manually conducted by human experts. Because instances are considered to an important evidence reflecting the class, We can regard that there might be certain relationships between both sets of instances I_i and I_j in the aligned classes c_i and c_j.

Thereby, some of terms in the instances should be extracted and regarded as principal components representing the class.

Definition 4 (*Term features*). Term features from a certain set of textual instances are extracted by using dimensionality reduction methods. Let T_i be a set of terms extracted from the preprocessed instances I_i in class c_i. Term features F_i^{term} is given by

$$F_i^{\mathrm{term}} = \{t_m | \mathcal{R}(t_m) \geqslant \tau_{\mathrm{related}}, t_m \in T_i\} \tag{3}$$

where function $\mathcal{R}$ indicates the rate of relatedness to c_i. It can be computed by statistical analysis for term occurrence patterns. As compared with a threshold τ_{related} determined

by a user, some term feature whose relatedness is higher than the others are selected.

There exist several well-known methods such as feature subset selection based on term weighting and Bayesian networks. We apply latent semantic analysis (LSA) (Deerwester, Dumais, Landauer, Furnas, & Harshman, 1990) using singular value decomposition (SVD) to the instances. In fact, a set of instances $I_p = \{d_1, d_2, \ldots, d_{|I_p|}\}$ in class c_p are represented as term-document matrix M_{TD} of which size is $|T_p| \times |I_p|$. Each element $td_{i,j}$ is computed by

$$td_{i,j} = TF_{i,j} \times \log \frac{1}{\mathrm{IDF}_i} = \frac{\mathrm{Occur}(t_i, d_j)}{|d_j|} \times \log \frac{|I_p|}{|\{d_k | \mathrm{Occur}(t_i, d_k) \geqslant 1, d_k \in I_p\}|} \quad (4)$$

where $\mathrm{Occur}(a,b)$ returns the number of occurrences of term a in document b. It is simply based on TF-IDF (term frequency and inverse document frequency) weights. Then, by SVD method, the matrix M_{TD} is decomposed into

$$M_{\mathrm{TD}} = UDV \quad (5)$$

where D is a diagonal matrix. We want to skip more description about SVD (see Golub & Van Loan, 1996). The ith diagonal component in D indicates the relatedness $\mathcal{R}(t_i)$ of term t_i. According to the Eq. (3), the highly ranked terms over threshold value τ_{related} become chosen.

In particular, we expect that some potential matchings between term features $\langle f_\alpha, rel, f_\beta \rangle^*$ can be implied by a set of manual alignments $\langle c_i, rel, c_j \rangle$ given by human experts. It means that with a given relation rel the pairs of features $f_\alpha \in F_i^{\mathrm{term}}$ and $f_\beta \in F_j^{\mathrm{term}}$ are obviously associated with each other. These information, most importantly, can be applied to make it more accurate to measure the similarities between neighbor classes.

3.2. Alignment based on class similarity

In order to find optimal alignment between two taxonomies, we have to measure the similarity between classes consisting of the taxonomies.

Definition 5 (*Class similarity*). Given a pair of classes from two different taxonomies, the class similarity (Sim_C) between c and c' is defined as

$$Sim_C(c, c') = \sum_{E \in \mathcal{N}(C)} \pi_E^C MSim_Y(E(c), E(c')) \quad (6)$$

where $\mathcal{N}(C) \subseteq \{E^1 \ldots E^n\}$ is the set of all relationships in which the classes participate (for instance, subclass, superclass, or instances). We have to consider on three components $Y = \{L, C, F\}$ (i) class labels (L), (ii) neighboring classes (C), and (iii) instead of instances, term features (F^{term}). The weights π_E^C are normalized (i.e., $\sum_{E \in \mathcal{N}(C)} \pi_E^C = 1$). Class similarity measure Sim_C is assigned in [0,1].

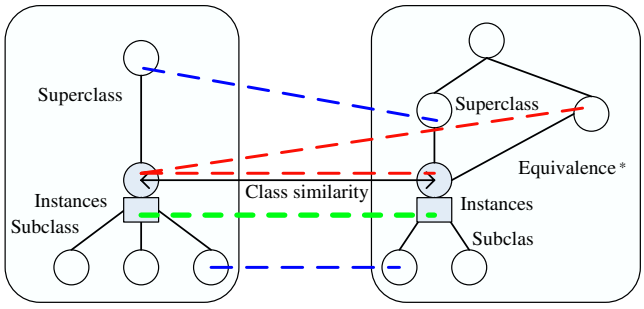

Fig. 3. Similarity-based business taxonomy alignment.

As a matter of fact, a similarity function between two set of classes can be established by finding a maximal matching maximizing the summed similarity between the classes:

$$MSim_C(S, S') = \frac{\max(\sum_{\langle c,c' \rangle \in \mathrm{Pairing}(S,S')}(Sim_C(c, c')))}{\max(|S|, |S'|)}, \quad (7)$$

in which *Pairing* provides a matching of the two set of classes. Methods like the Hungarian method allow to find directly the pairing which maximizes similarity. The algorithm is an iterative algorithm that compute this similarity (Euzenat & Valtchev, 2004). This measure is normalized because if Sim_C is normalized, the divisor is always greater or equal to the dividend.

In case of business taxonomies, according to Definition 3, we have to take in account all possible relationships (*rel* and *rel**) between classes for $\mathcal{N}(C) = \{E^{\mathrm{sup}}, E^{\mathrm{sub}}, E^{\mathrm{equ}}\}$, provided (i) the superclass and the subclass defined in each faceted taxonomy (depicted as blue dotted lines in Fig. 3), and (ii) E^{sup}, E^{sub}, the equivalent class (E^{equ}) by manual alignments of human experts (depicted as red dotted lines in Fig. 3), respectively.[2] Then, Eq. (6) can be rewritten as:

$$\begin{aligned} Sim_C(c, c') = {} & \pi_L^C sim_L(L(c), L(c')) \\ & + \pi_{\mathrm{sub}}^C MSim_C(E^{\mathrm{sub}}(c), E^{\mathrm{sub}}(c')) \\ & + \pi_{\mathrm{sup}}^C MSim_C(E^{\mathrm{sup}}(c), E^{\mathrm{sup}}(c')) \\ & + \pi_{\mathrm{equ}}^C MSim_C(E^{\mathrm{equ}}(c), E^{\mathrm{equ}}(c')) \\ & + \pi_{F^{\mathrm{term}}}^C sim_{F^{\mathrm{term}}}(F_c^{\mathrm{term}}, F_{c'}^{\mathrm{term}}) \end{aligned} \quad (8)$$

where the set functions $MSim_C$ compute the similarity of two entity collections. Label similarity sim_L is simply computed by string matching algorithms such as *Levenshtein* edit distance (Levenshtein, 1996), substring distance (Euzenat, 2004), and so on. Similarity measure between two classes can be turned into a distance measure Distance = 1 − Similarity by taking its complement to 1.

Especially, in order to enhance the accuracy of the class similarity, the last term in Eq. (8) is representing

[2] For interpretation of color in Fig. 3, the reader is referred to the web version of this article.

instance-level similarity measurement between the term features extracted from the instances (shown as a green dotted line in Fig. 3). We exploit three different heuristic functions, and they are formulated by

$$Sim_{F^{\text{term}}}(F_c^{\text{term}}, F_{c'}^{\text{term}}) = \frac{N}{\max(|F_c^{\text{term}}|, |F_{c'}^{\text{term}}|)} \tag{9}$$

$$= \max_{n=1}^{N} Sim_{\langle f_\alpha, f_\beta\rangle \in Pairing(F_c^{\text{term}}, F_{c'}^{\text{term}})}(L(f_\alpha), L(f_\beta))_n \tag{10}$$

$$= \frac{\sum_{n=1}^{N} Sim_{\langle f_\alpha, f_\beta\rangle \in Pairing(F_c^{\text{term}}, F_{c'}^{\text{term}})}(L(f_\alpha), L(f_\beta))_n}{N} \tag{11}$$

where N is the number of pairs of term features whose distances computed by string matching methods are less than threshold τ_{Dist}, e.g., EditDistance$(L(f_\alpha), L(f_\beta)) \leqslant \tau_{\text{Dist}}$. Three equations are denoted as H_1, H_2, and H_3, and they return the normalized number of matched pairs of terms, the maximum similarity among matched terms, and the average similarity of matched terms, respectively. (We will evaluate and compare these heuristic functions for matching term features in Section 6.) Assuming that instance-level class similarity can uncover the latent semantic information of the classes, the normalization process is expected to prune incorrect alignments between them, by comparing threshold τ_{Align}.

Hence, the alignments between heterogeneous taxonomies can be represented as a set of pairs of concepts from two different taxonomies. We refer these concept pairs to correspondences (e.g., equivalence or subsumption).

Definition 6 (*Alignment*). Given two taxonomies $\mathcal{T}_i$ and $\mathcal{T}_j$, the alignments between two taxonomies are represented as a set of correspondences $\text{CRSP}_{ij} = \{\langle c, rel, c'\rangle | c \in \mathcal{T}_i, c' \in \mathcal{T}_j\}$ where *rel* means the relationship between c and c', by maximizing the summation of class similarities $\sum Sim_{\mathcal{C}(c,c')}$.

Finally, alignment process makes heterogeneous VEs interoperable (even partially) among them. For example, local users in a VE can easily and transparently access to the other VEs. To do so, VEs have to conduct the taxonomy alignment process in advance. Suppose that a set of VEs $\{L_1, \ldots, L_N\}$ should be interoperable with each other. Alignment process can find out the correspondences between all pairs of taxonomies, i.e., L_i obtains $N-1$ sets of correspondences.

3.2.1. Example

We want to show a simple example. In Fig. 4, the alignment between two taxonomies is occurred as showing the best mappings between them. All relations mean only subclass relations, and the weights are assumed as $\pi_L^C = 0.8$, $\pi_{\text{sub}}^C = 0.2$. In the alignment between taxonomy $\mathcal{T}_1$ and $\mathcal{T}_2$, label similarity based on *Levenshtein* edit distance is

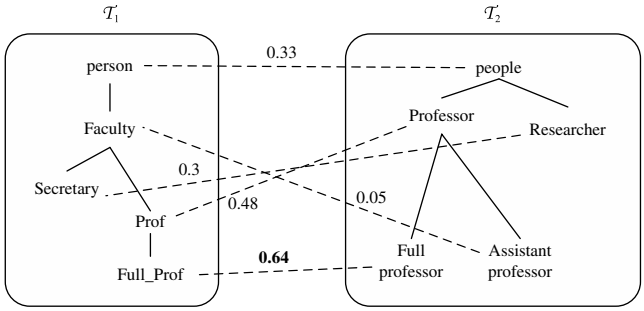

Fig. 4. Example of similarity-based alignment.

measured by $sim_L(c_i, c_j) = 1 - \frac{\text{Dist}(c_i, c_j)}{\max(|c_i|, |c_j|)}$. The maximal label similarity of pairs of classes is calculated as follows.

$$sim_L(\text{person}, \text{Professor}) = 0.44 \tag{12}$$

$$sim_L(\text{Faculty}, \text{people}) = 0.14$$

$$sim_L(\text{Secretary}, \text{Researcher}) = 0.30$$

$$sim_L(\text{Prof}, \text{Professor}) = 0.44 \tag{13}$$

$$sim_L(\text{Full\_Prof}, \text{Full\_professor}) = 0.64, \tag{14}$$

Then, the most similar pair of classes can be found out 'Full_Prof' of $\mathcal{T}_1$ and 'Full professor' of $\mathcal{T}_2$ in Eq. (14). As shown in Eqs. (12) and (13), classes 'person' and 'Prof' in $\mathcal{T}_1$ have shown the same label similarities with 'Professor' in $\mathcal{T}_2$. We have to pay attention to the subclasses. Thus, the summation of similarities are calculated by

$$Sim_{\mathcal{C}}(\text{person}, \text{Professor}) = 0.8 \times 0.44 + 0.2 \times \max(0.07, 0.05) = 0.366 \tag{15}$$

where $Sim_{\mathcal{C}}(\text{Faculty}, \text{Full\_professor})$ and $Sim_{\mathcal{C}}(\text{Faculty}, \text{Assistant\_professor})$ are assigned into 0.07 are 0.05, respectively. Similarly, we can calculate

$$Sim_{\mathcal{C}}(\text{Prof}, \text{Professor}) = 0.352 + 0.2 \times \max(0.64, 0.26) = 0.48, \tag{16}$$

where $Sim_{\mathcal{C}}(\text{Full\_Prof}, \text{Full\_professor})$ and $Sim_{\mathcal{C}}(\text{Full\_Prof}, \text{Assistant\_professor})$ are 0.64 and 0.26, respectively. This means that 'Prof' in $\mathcal{T}_1$ has to be aligned to 'Professor' in $\mathcal{T}_2$. Moreover, we employ a threshold τ_{Align} to filter out some correspondences of which class similarities are less than the threshold.

In addition, the term features extracted from instances should be considered. Suppose that two classes 'Prof' $\in \mathcal{T}_1$ and 'Professor' $\in \mathcal{T}_2$ are manually aligned with arbitrary other classes in the same taxonomies. Then, two set of term features extracted from these two classes are given by

- $F_{\text{Prof}}^{\text{term}} = \{\text{professor}, \text{teach}, \text{university}, \text{lecture}\}$, and
- $F_{\text{Professor}}^{\text{term}} = \{\text{universe}, \text{lecture}, \text{position}, \text{doctor}, \text{profess}\}$, respectively.

We can compute the similarities between all possible pairs of terms from $F_{\text{Prof}}^{\text{term}}$ and $F_{\text{Professor}}^{\text{term}}$, as shown in Table 1. If

$\tau_{\text{Dist}} = 0.5$, we found out that $N = |\{\langle universe, university\rangle, \langle lecture,lecture\rangle, \langle professor,profess\rangle\}| = 3$. Based on three heuristics, the $Sim_{F^{\text{term}}}$ in Eq. (8) can be computed by

$$\begin{aligned} Sim_{F^{\text{term}}}(\text{Prof}, \text{Professor}) &= \frac{3}{5} = 0.6 = \max[0.7, 1.0, 0.778] \\ &= 1.0 = \frac{0.7 + 1.0 + 0.778}{3} \\ &= 0.826 \end{aligned}$$

respectively. These results can reinforce the class similarity in Eq. (16) measured by only semantic structure of taxonomies. If H_3 is chosen, Eq. (16) is extended to

$$\begin{aligned} Sim_C(\text{Prof}, \text{Professor}) &= 0.6 \times 0.44 + 0.2 \times 0.64 \\ &\quad + 0.2 \times 0.826 \\ &= 0.58 \end{aligned} \tag{17}$$

where $\pi_L^C = 0.6$, $\pi_{\text{sub}}^C = 0.2$ and $\pi_{F^{\text{term}}}^C = 0.2$.

After normalization, we can expect to remove some mismatched alignments (e.g., between 'Secretary' and 'Researcher') by decreasing their class similarities.

4. Interoperability based on query transformation

Each VE can interact with others by using the correspondences obtained from taxonomy alignment process. If their interactions are simply based on (i) query answering and (ii) recommending (in other words, pushing) tasks for relevant information exchanging, we focus on conceptual transformation of the queries which indicate some specific information needs of the VEs.

During communicating between VEs, the queries can be embedded into the messages sent from a source VE to a destination VE.

Definition 7 (*Query*). A query Q is composed of a set of classes (or terms) and logical operators (e.g., $\neg$, $\wedge$, and $\vee$), and its grammar is simply given by

$$q ::= c | \neg q | q \wedge q | q \vee q \tag{18}$$

where $c \in \mathcal{T}_{src}$ of the source VE.

For conceptual query transformation from VE_i to VE_j, we exploit simple class replacement strategy using a set of aligned correspondences CRSP_{ij} between $\mathcal{T}_i$ and $\mathcal{T}_j$, in order to enhance the accessibility of proactive software modules (e.g., agents) and, more particularly, local users.

Table 1
Similarities between term features for class similarity ($\tau_{\text{Dist}} = 0.5$)

	Professor	Teach	University	Lecture
Universe	0.222	0.125	**0.7**	0.25
Lecture	0.111	0.143	0.1	**1.0**
Position	0.333	0.0	0.1	0.125
Doctor	0.333	0.0	0.1	0.429
Profess	**0.778**	0.0	0.2	0.0

In other words, we want to help a local user in VE_i to search for relevant resources in heterogeneous VEs by replacing the concepts in queries.

Definition 8 (Query transformation). Let a query q_i in VE_i be sent to VE_j, and divided into

$$q_i = q_{ij}^+ + q_{ij}^- = \{c^+ | \langle c^+, rel, c'\rangle \in \text{CRSP}_{ij}\} + \{c^- | c^- \in \mathcal{T}_i\} \tag{19}$$

where class c^+ is matched with a certain class in $\mathcal{T}_j$. This query is transformed by replacing class c^+ in q_i with the classes c' in $\mathcal{T}_j$, if and if only

- c' is equivalent with c^+ (rel = Equivalence), or
- c' is a subclass of c^+ (rel = SubClass).

As an example, in VE_i, a query "Media $\wedge$ Art" expresses the intersection between two sets of resources annotated with classes "Media" and "Art," respectively. If a correspondence $\langle$Media,SubClass,Video$\rangle$ is discovered between $\mathcal{T}_i$ and $\mathcal{T}_j$, the query can be modified to "Video $\wedge$ Art" in VE_j.

In case of replacement with subclasses, the transformed queries are expressing more specified concepts. In terms of recall and precision (well-known measurements from information retrieval field), it makes the precision of the retrieved information more increased, while the coverage rate (or recall) is reduced. On the other hand, query transformation based on superclass replacement may cause information loss problem, because the transformed query is impossible to indicate the specific semantics of the original one.

5. Experimental results

We have evaluated our contributions of this paper by two main issues; (i) human evaluation of alignment between heterogeneous taxonomies, and (ii) performance evaluation (i.e., recall and precision) of resource retrieval based on query transformation.

Above all, in order to prepare a testing bed, we have invited 12 students, and asked them to their own VEs with respect to their preferences. Given a set of resources (i.e., product description files[3]), they had to choose a number of resources to annotate with their taxonomies. They were able to merge faceted taxonomies, and assert manual alignments between these merged faceted taxonomies. The faceted taxonomies were simply obtained by screening some parts of existing taxonomies. Such taxonomies are

- ACM Computing Classification (http://www.acm.org/class/1998/)
- Government Category List (http://www.esd.org.uk/standards/gcl/)

[3] IRCS dataset is organized as a set of documents retrieved from major e-commerce websites in Korea (available on http://eslab.inha.ac.kr/~ircs/). It has been applied to support user browsing tasks for searching relevant information in (Jung, 2007a).

Table 2
Specifications of testing bed

	Number of resources	Number of classes ($\lvert\mathcal{T}_i\rvert$)	Density ($\frac{\text{Number of resources}}{\text{Number of classes}}$)
VE_1	172	37	Middle (4.65)
VE_2	81	25	Low (3.24)
VE_3	59	18	Low (3.28)
VE_4	73	27	Low (2.70)
VE_5	**614**	**57**	High (10.77)
VE_6	264	21	High (12.57)
VE_7	**510**	48	High (10.63)
VE_8	236	29	Middle (8.14)
VE_9	69	16	Middle (4.31)
VE_{10}	276	**60**	Middle (4.60)
VE_{11}	185	23	Middle (8.04)
VE_{12}	243	28	Middle (8.68)

- On-line Medical Dictionary (OMD) (http://cancerweb.ncl.ac.uk/omd/).
- Open Directory Project (ODP) (http://dmoz.org/), and
- Commerce-Database Business Directory (http://www.commerce-database.com/).

Table 2 shows the specifications of our testing bed. While VE_5 and VE_7 have annotated the largest number of resources, VE_5 and VE_{10} have shown the largest number classes in the corresponding taxonomies. With respect to the density ($D = \frac{\text{Numberofresources}}{\text{Numberofclasses}}$), the VEs are classified into three categories; high, middle, and low. Particularly, VE_6 was designed to be the densest one ($D_6 = 12.57$).

5.1. Evaluation on alignment process

For the first issue, we performed alignment process between all possible pairs of taxonomies $\left(\frac{12\times11}{2} = 66\right)$ in three difference cases; (i) simple matching of semantic structures of taxonomies, (ii) matching with manual alignments $\mathcal{A}$, and (iii) matching with term features extracted from instance sets. Compared with the matching result in the first case, second and third cases were expected to show improved results. Five human experts, thereby, analyzed the collected correspondences between taxonomies, and counted the number of mismatched correspondences from each alignment.

Table 3 shows the results of correspondence matched in the first case. While lower diagonal component $v_{ij} = |\text{CRSP}_{ij}|$ is the number of correspondences between two taxonomies $\mathcal{T}_i$ and $\mathcal{T}_j$, upper diagonal component w_{ji} mean the number of mismatched correspondences from CRSP_{ij}. Additionally, in the bracket, mismatching ratio is computed by $\frac{v_{ij}}{w_{ji}}$. In average, our similarity-based taxonomy alignment has shown approximately 29% mismatching ratio. Taxonomy alignment between $\mathcal{T}_1$ and $\mathcal{T}_5$ has shown 44%, which is the highest mismatching ratio (i.e., the worst case). On the other hand, alignments between $\mathcal{T}_3$ and $\mathcal{T}_5$, between $\mathcal{T}_3$ and $\mathcal{T}_{10}$ were the lowest mismatching ratio 14% (i.e., the best case).

In order to enhance the previous alignments, we exploited two approaches; (i) manual alignments $\mathcal{A}$, provided by the students during building their VEs, and (ii) term features representing each class, extracted from instances of the corresponding class. From three heuristic functions (Eq. (9)–(11)), we chose first funtion (H_1). Tables 4 and 5 show the alignment results in both cases (Similar to the Table 3, lower and upper diagonal components in these tables indicate the numbers of correspondences and mismatched ones by human experts, respectively).

We want to evaluate whether (and how much) these methods were able to improve the alignment performance. Thereby, we compared the experimental results in the previous tables. With respect to improvement of the number of discovered correspondences, as shown in Fig. 5, instance-based alignment (about 139%) outperformed manual alignments-based one (about 105%). We found out that in most of taxonomy pairs the manual alignment has shown only slight improvement, compared to the instance-based matching. Particularly, although the ratio of manual alignment $\rho_{\text{Align}}(k)$ in the twelve taxonomies was diverse between $\rho_{\text{Align}}(10) = 0.32$ and $\rho_{\text{Align}}(9) = 0.65$, their performance was quite consistently maintained. It means that manual alignment has played trivial contributions to automatic taxonomy alignment.

With respect to the mismatching ratio, as shown in Fig. 6, two methods decreased, in average, 9.8% and 48.4% of mismatched correspondences, which are regarded as error rates. Again, instance-based alignment has shown better

Table 3
Results of taxonomy alignment ($\tau_{\text{Align}} = 0.1$)

	$\mathcal{T}_1$	$\mathcal{T}_2$	$\mathcal{T}_3$	$\mathcal{T}_4$	$\mathcal{T}_5$	$\mathcal{T}_6$	$\mathcal{T}_7$	$\mathcal{T}_8$	$\mathcal{T}_9$	$\mathcal{T}_{10}$	$\mathcal{T}_{11}$	$\mathcal{T}_{12}$
$\mathcal{T}_1$	–	2	2	4	8 (**0.44**)	3	6	4	2	5	3	5
$\mathcal{T}_2$	11	–	3	5	5	3	2	4	2	4	4	5
$\mathcal{T}_3$	9	8	–	1	1 (**0.14**)	2	2	2	1	1 (**0.14**)	3	2
$\mathcal{T}_4$	15	13	6	–	5	4	3	4	2	3	4	3
$\mathcal{T}_5$	18	13	7	15	–	3	6	4	2	**9**	5	3
$\mathcal{T}_6$	9	8	8	11	10	–	3	4	1	4	4	3
$\mathcal{T}_7$	19	11	8	13	16	11	–	3	3	8	3	5
$\mathcal{T}_8$	15	11	6	15	15	10	12	–	2	3	2	3
$\mathcal{T}_9$	6	9	6	7	7	5	8	7	–	2	3	2
$\mathcal{T}_{10}$	13	11	7	9	**30**	10	25	12	7	–	3	3
$\mathcal{T}_{11}$	9	12	10	13	12	11	10	8	9	11	–	4
$\mathcal{T}_{12}$	14	13	6	11	15	7	15	13	7	10	10	–

Table 4
Results of taxonomy alignment with manual alignment ($\tau_{\text{Align}} = 0.1$)

	$\mathcal{T}_1$	$\mathcal{T}_2$	$\mathcal{T}_3$	$\mathcal{T}_4$	$\mathcal{T}_5$	$\mathcal{T}_6$	$\mathcal{T}_7$	$\mathcal{T}_8$	$\mathcal{T}_9$	$\mathcal{T}_{10}$	$\mathcal{T}_{11}$	$\mathcal{T}_{12}$
$\mathcal{T}_1$	–	2	2	3	7	3	5	4	2	4	3	4
$\mathcal{T}_2$	12	–	2	5	5	3	2	4	2	3	3	4
$\mathcal{T}_3$	9	8	–	1	1	2	2	2	1	1	3	2
$\mathcal{T}_4$	16	14	6	–	5	4	3	3	2	3	4	2
$\mathcal{T}_5$	19	14	7	16	–	2	5	4	2	**9**	5	3
$\mathcal{T}_6$	9	9	8	11	11	–	3	3	1	4	3	3
$\mathcal{T}_7$	20	11	8	14	17	11	–	3	3	8	3	5
$\mathcal{T}_8$	16	11	6	16	16	10	13	–	2	2	2	3
$\mathcal{T}_9$	6	9	6	7	7	5	8	7	–	2	3	2
$\mathcal{T}_{10}$	13	12	7	9	**32**	11	27	12	7	–	2	3
$\mathcal{T}_{11}$	9	13	10	14	12	11	10	8	10	11	–	3
$\mathcal{T}_{12}$	14	14	6	11	16	7	16	14	7	11	10	–

Table 5
Results of taxonomy alignment with term features extracted from instances ($\tau_{\text{Align}} = 0.1$)

	$\mathcal{T}_1$	$\mathcal{T}_2$	$\mathcal{T}_3$	$\mathcal{T}_4$	$\mathcal{T}_5$	$\mathcal{T}_6$	$\mathcal{T}_7$	$\mathcal{T}_8$	$\mathcal{T}_9$	$\mathcal{T}_{10}$	$\mathcal{T}_{11}$	$\mathcal{T}_{12}$
$\mathcal{T}_1$	–	1	1	2	4	1	4	1	1	2	2	3
$\mathcal{T}_2$	16	–	2	2	3	2	1	1	1	2	2	2
$\mathcal{T}_3$	11	10	–	**0**	1	1	1	1	1	**0**	2	1
$\mathcal{T}_4$	19	19	8	–	2	3	2	3	1	2	1	2
$\mathcal{T}_5$	25	17	11	22	–	1	3	2	1	4	2	2
$\mathcal{T}_6$	14	11	12	17	12	–	2	3	**0**	3	2	2
$\mathcal{T}_7$	25	14	11	19	20	16	–	2	1	3	2	3
$\mathcal{T}_8$	22	15	9	22	22	14	19	–	1	2	1	1
$\mathcal{T}_9$	8	12	8	9	8	6	12	9	–	1	2	1
$\mathcal{T}_{10}$	17	16	10	11	39	16	33	17	9	–	1	2
$\mathcal{T}_{11}$	13	19	13	21	15	17	13	12	11	15	–	2
$\mathcal{T}_{12}$	18	17	9	16	23	11	23	19	9	14	14	–

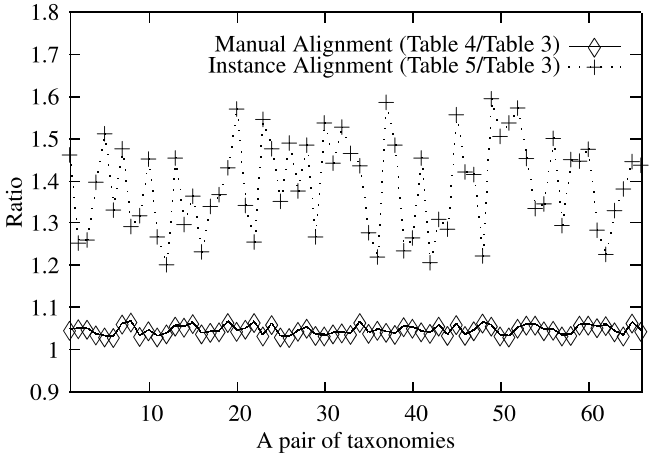

Fig. 5. Ratio of number of discovered correspondences.

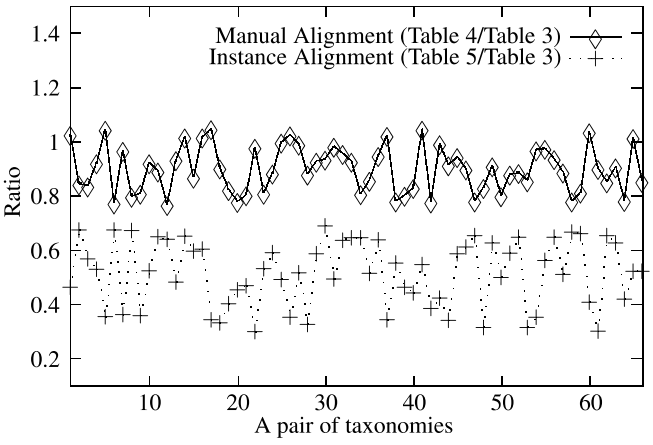

Fig. 6. Ratio of number of mismatched correspondences.

performance than others. Especially, instance-based alignments between taxonomies $\mathcal{T}_3$ and $\mathcal{T}_4$, between $\mathcal{T}_3$ and $\mathcal{T}_{10}$, and between $\mathcal{T}_6$ and $\mathcal{T}_9$ have been perfectly matched.

5.2. Evaluation on query transformation

Second experimentation issue is to evaluate semantic interoperability between VEs. In this paper, interactions between VEs were represented as concept-based queries, and these queries were transformed by class replacement based on the correspondences, acquired by instance-level alignment method in the first issue. The invited students have built ten queries with the classes in their own taxonomies, in order to broadcast these queries to the rest of VEs. After a set of resources $\widetilde{rsc}(q_i)$ were retrieved by a query q_i of VE_i, the recall R and precision P have been measured by

$$R(q_i) = \frac{|\widetilde{rsc}(q_i) \cap rsc(q_i)|}{|rsc(q_i)|} \tag{20}$$

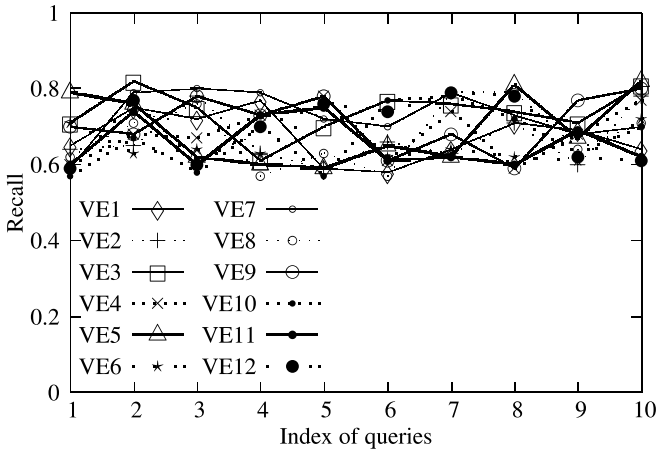

Fig. 7. Recall measurement for query transformation.

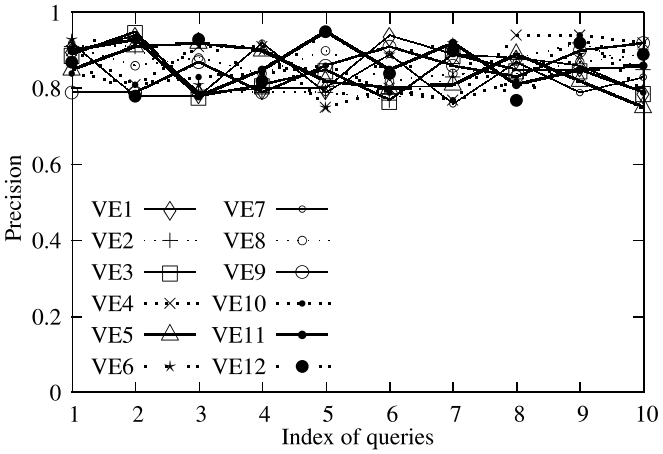

Fig. 8. Precision measurement for query transformation.

$$P(q_i) = \frac{|\widetilde{rsc}(q_i) \cap rsc(q_i)|}{|\widetilde{rsc}(q_i)|} \tag{21}$$

where $rsc(q_i)$ is a set of resources retrieved by the human experts.

For the given queries, Figs. 7 and 8 are showing the experimental results of both measurements, respectively. Average recall was 68.6%, and the queries from VE_3 have been most successfully transformed (73.7%). With respect to precision, we obtained in average 85.4% precision. VE_{11} has shown the maximum precision (86.9%).

It proves that the correspondences were properly discovered by the proposed approach (i.e., the rate of mismatched alignments is reasonably low), but some missed correspondences made the recall decreased. Another important point is that precision measure has shown better results rather than recall measure. We consider that our concept replacement strategy is only based on "equivalence" and "subclass" relationships.

6. Discussion and related work

Through conducting experimentation, the proposed alignment has been proved to support semantic interoperability between heterogeneous VEs. We want to discuss several meaningful achievements related to taxonomy alignment algorithm.

First issue is to find out whether the characteristics of VEs (e.g., numbers of resources and classes in Table 2) are related to the performance of alignment process or not. Given two taxonomies $\mathcal{T}_i$, $\mathcal{T}_j$ to be aligned, four parameters were chosen to be compared, as follows:

- Total numbers of two sets of classes (p_1^{ij})
- Total numbers of two sets of instances (p_2^{ij})
- Difference between numbers of two sets of classes (p_3^{ij})
- Difference between numbers of two sets of instances (p_4^{ij})

Then, some meaningful associations between two known quantitative variables, i.e., a parameter p_k and improved ratio of alignment process r, has been analyzed by regression method

$$r^{ij} = \alpha + \beta \times p_k^{ij} + \gamma \times (p_k^{ij})^2 + \epsilon \tag{22}$$

where α, β, and γ are coefficients. We found out that the larger difference between the numbers of classes in taxonomies (i.e., p_3) make the best influence on the performance of taxonomy alignment.

As second issue, we found out our alignment process has shown approximately 35.5% error rate (i.e., the mismatched correspondences). In the worst case (alignment between $\mathcal{T}_1$ and $\mathcal{T}_2$), we realized that mainly the differences between domain-specific terminologies have influenced string matching-based alignments (in our case, we measured the edit distance between labels).

Our approach can be compared with the centralized information systems, e.g., portal system. Difference between two main approaches to access to multiple information sources, in terms of end-users' accessing strategies. While portal systems (e.g., meta search engines) provide a centralized integration service from these information sources, distributed approaches like our system can consider more domain-specific features. Moreover, they can expect some personalization techniques to their local users.

We consider the taxonomies are a part of ontologies in semantically heterogeneous environment. While the main relationship between classes in taxonomies is `SubClass`, ontologies are containing a variety of relationships between classes such as `SubClass`, `SuperClass`, `Property`, `SubProperty`, `Domain`, `Range`, and so on. However, in Welty and Guarino (2001), the taxonomic patterns are capable of ontological relationships. Also, many work has been proposed to match, align and merge taxonomies like similarity flooding (Melnik, Garcia-Molina, & Rahm, 2002), Alignment API (Ehrig & Sure, 2005) and directory-based approach (Liang, Vaishnavi, & Vandenberg, 2006). Of particular interest is ontology sharing system between community of practice (cop), introduced in Davies, Duke, and Sure (2004) Mika, Iosif, Sure, and Akkermans (2004). In more practical aspect, several busi-

ness markup languages, e.g., Unified Enterprise Modelling Language (UEML) (Ducq, Chen, & Vallespir, 2004), have been designed. It can be regarded as more systematic activities.

In context of query transformation, since concept-based query transformation scheme was introduced in Qiu and Frei (1993), several approaches have been investigated. Examples of such approaches are probabilistic query expansion based on concept similarity (Cui, Wen, Nie, & Ma, 2002), logical inference (Nie, 2003), and background knowledge-based systems (Liu & Chu, 2005; Zazo, Figuerola, Alonso Berrocal, & Rodríguez, 2005).

7. Concluding remarks and future work

As a conclusion, we proposed alignment-based query transformation scheme on heterogeneous virtual organizations (in particular, virtual enterprises). Each pair of taxonomies were aligned by measuring the similarities between classes. We assume that the maximal summation of these class similarities be the best alignment between the corresponding taxonomies. Based on this alignment, we supported the local users to access to the other heterogeneous VEs.

In the future, we have to evaluate the scalability of our alignment-based distributed VEs, as increasing the number of testing beds. Especially, according to the semantic power, they might be socialized, as shown in Jung and Euzenat (2006) and Jung (2007b). Then, we can provide more efficient query propagation strategies. More importantly, we are planning to evaluate our alignment method by evaluation methods of taxonomy mapping algorithms proposed in Avesani, Giunchiglia, and Yatskevich (2005).

References

Abrol, Mani, Doshi, Bhavin, Kanihan, Jim, Kumar, Amit, Liu, Jinhui, & Mao, Jianchang (2005). Intelligent taxonomy management tools for enterprise content. In Andrzej Skowron, Rakesh Agrawal, Michael Luck, Takahira Yamaguchi, Pierre Morizet-Mahoudeaux, & Jiming Liu, et al. (Eds.), *Proceedings of the 2005 IEEE/WIC/ACM international conference on web intelligence (WI 2005)* (pp. 809–811). IEEE Computer Society.

Avesani, Paolo, Giunchiglia, Fausto, & Yatskevich, Mikalai (2005). A large scale taxonomy mapping evaluation. In Yolanda Gil, Enrico Motta, V. Richard Benjamins, & Mark A. Musen (Eds.), *International semantic web conference. Lecture Notes in Computer Science* (Vol. 3729, pp. 67–81). Springer.

Lopes Cardoso, Henrique, & Oliveira, Eugénio C. (2004). Virtual enterprise normative framework within electronic institutions. In Marie Pierre Gleizes, Andrea Omicini, & Franco Zambonelli (Eds.), *Proceedings of the 5th international workshop on engineering societies in the agents World (ESAW 2004). Lecture Notes in Computer Science* (Vol. 3451, pp. 14–32). Springer.

Cilia, M., & Buchmann, A. P. (2002). An active functionality service for e-business applications. *SIGMOD Records, 31*(1), 24–30.

Deerwester, Scott C., Dumais, Susan T., Landauer, Thomas K., Furnas, George W., & Harshman, Richard A. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science, 41*(6), 391–407.

Euzenat, Jérôme (2004). An API for ontology alignment. In Sheila A. McIlraith, Dimitris Plexousakis, & Frank van Harmelen (Eds.), *Proceedings of the 3rd international semantic web conference. Lecture Notes in Computer Science* (Vol. 3298, pp. 698–712). Springer.

Menczer, Filippo (2004). Lexical and semantic clustering by web links. *Journal of the American Society for Information Science and Technology, 55*(14), 1261–1269.

Golub, G. H., & Van Loan, C. F. (1996). *Matrix computations* (third ed.). Johns Hopkins University Press.

Guido, Vetere, & Maurizio, Lenzerini (2005). Models for semantic interoperability in service-oriented architectures. *IBM Systems Journal, 44*(4), 887–903.

Cui, Hang, Wen, Ji-Rong, Nie, Jian-Yun, & Ma, Wei-Ying (2002). Probabilistic query expansion using query logs. In *Proceedings of the 11th international conference on World Wide Web* (pp. 325–332). New York, NY, USA: ACM Press.

Jonkers, Henk, Lankhorst, Marc M., van Buuren, René, Stijn, Hoppenbrouwers, Bonsangue, Marcello M., & van der Torre, Leendert W. N. (2004). Concepts for modeling enterprise architectures. *International Journal of Cooperative Information Systems, 13*(3), 257–287.

Hull, Richard (1997). Managing semantic heterogeneity in databases: a theoretical prospective. In *Proceedings of the sixteenth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems (PODS 97)* (pp. 51–61). New York, NY, USA: ACM Press.

Euzenat, Jérôme, & Valtchev, Petko (2004). Similarity-based ontology alignment in OWL-Lite. In Ramon López de Mántaras & Lorenza Saitta (Eds.), *Proceedings of the 16th European conference on artificial intelligence (ECAI'2004)* (pp. 333–337). Valencia, Spain: IOS Press. August 22–27.

Nie, Jian-Yun (2003). Query expansion and query translation as logical inference. *Journal of the American Society for Information Science and Technology, 54*(4), 335–346.

Davies, John, Duke, Alistair, & Sure, York (2004). OntoShare – An ontology-based knowledge sharing system for virtual communities of practice. *Journal of Universal Computer Science, 10*(3), 262–283.

Jung, Jason J. (2005). Collaborative web browsing based on semantic extraction of user interests with bookmarks. *Journal of Universal Computer Science, 11*(2), 213–228.

Jung, Jason J. (2006). Taxonomy alignment for interoperability between heterogeneous digital libraries. In *Proceedings of the International Conference on Asian Digital Library (ICADL). Lecture Notes in Computer Science* (Vol. 4312, pp. 274–282). Springer.

Jung Jason, J. (2007a). Exploiting semantic annotation to supporting user browsing on the web. *Knowledge-Based Systems, 20*(4), 373–381.

Jung Jason, J., (2007b). Ontological framework based on contextual mediation for collaborative information retrieval. *Information Retrieval, 10*(1), 85–109.

Jung Jason J. & Euzenat Jérôme (2006). From personal ontologies to semantic social space. In *Poster of the 4th European Semantic Web Conference (ESWC 2006)*.

Levenshtein, I. V. (1996). Binary codes capable of correcting deletions, insertions, and reversals. *Cybernetics and Control Theory, 10*(8), 707–710.

Liang, Jianghua, Vaishnavi, Vijay K., & Vandenberg, Art (2006). Clustering of LDAP directory schemas to facilitate information resources interoperability across organizations. *IEEE Transactions on Systems, Man, and Cybernetics – Part A, 36*(4), 631–642.

Liu, Zhenyu, & Chu, Wesley W. (2005). Knowledge-based query expansion to support scenario-specific retrieval of medical free text. In *Proceedings of the 2005 ACM symposium on applied computing (SAC '05)* (pp. 1076–1083). New York, NY, USA: ACM Press.

Marc Ehrig & York Sure (2005). FOAM – framework for ontology alignment and mapping – results of the ontology alignment evaluation initiative. In Benjamin Ashpole, Marc Ehrig, Jérôme Euzenat, and Heiner Stuckenschmidt (Eds.), *Proceedings of the K-CAP 2005 workshop on integrating ontologies, Banff, Canada, October 2, 2005*, Vol. 156, CEUR Workshop Proceedings, CEUR-WS.org.

Shvaiko, Pavel, & Euzenat, Jérôme (2005). A survey of schema-based matching approaches. *Journal of Data Semantics, 4*, 146–171.

Mika, Peter, Iosif, Victor, Sure, York, & Akkermans, Hans (2004). Ontology-based content management in a virtual organization. In Steffen Staab & Rudi Studer (Eds.), *Handbook on Ontologies International Handbooks on Information Systems* (pp. 455–476). Springer.

Melnik, Sergey, Garcia-Molina, Hector, & Rahm, Erhard (2002). Similarity flooding: A versatile graph matching algorithm and its application to schema matching. In *Proceedings of the 18th International Conference on Data Engineering (ICDE)* (pp. 117–128). IEEE Computer Society.

Castano, Silvana, Ferrara, Alfio, & Montanelli, Stefano (2006). Matching ontologies in open networked systems: Techniques and applications. *Journal of Data Semantics, 5*, 25–63.

Welty, Christopher A., & Guarino, Nicola (2001). Supporting ontological analysis of taxonomic relationships. *Data & Knowledge Engineering, 39*(1), 51–74.

Qiu, Yonggang, & Frei, Hans-Peter (1993). Concept based query expansion. In *Proceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '93)* (pp. 160–169). New York, NY, USA: ACM Press.

Ducq, Yves, Chen, David, & Vallespir, Bruno (2004). Interoperability in enterprise modelling: requirements and roadmap. *Advanced Engineering Informatics, 18*(4), 193–203.

Zazo, Ángel F., Figuerola, Carlos G., Alonso Berrocal, José L., & Rodríguez, Emilio (2005). Reformulation of queries using similarity thesauri. *Information Processing and Management: An International Journal, 41*(5), 1163–1173.