14 ONTOLOGY ENGINEERING IN VIRTUAL BREEDING ENVIRONMENTS

Dora Simões[1,3], Hugo Ferreira[1], António Lucas Soares[1,2]

INESC Porto[1], Faculty of Engineering Univ. Porto[2], ISCA Univ. Aveiro[3]
dora.simoes@isca.ua.pt, hmf@inescporto.pt, asoares@inescporto.pt
PORTUGAL

This paper proposes a new method for managing the use of ontologies in the context of a Virtual Breeding Environment. This research work focus is on the dissolution phase of a Virtual Enterprise or Collaborative Network, where ontology segmentation techniques are user to enrich the VBE's ontology library. Firstly, an overview of the process of ontology composition and decomposition is given and the ontology library system adopted described. Then, the ontologies' ranking and classification method is described, explaining a set of metrics inspired in social network approaches. Finally, the results of preliminary tests are discussed.

1 INTRODUCTION

Current industrial and technological trends point to the need of SMEs to be able to enter in one or more networks in order to have access to a broader range of market opportunities. The long-term relational structure of cooperating SMEs and the operationalization of market opportunities by collaboration have been conceptualized in some scientific literature as Virtual Breeding Environments (VBE) and Virtual Enterprises (VE), respectively (Bacquet, Fatelnig et al., 2004; Afsarmanesh and Camarinha-Matos, 2005). A VBE is characterised by long-term goal relationships, establishing mutual trust between its members. This is the ideal context for the creation, operation and dissolution of VEs. This process should be as much as possible focused and efficient, in a way that cost and business effectiveness can be achieved. On the other hand we can generalise and consider a VBE not only as being a catalyst of VEs but also of Collaborative Networks (CN) (seen here as a special case of networks whose nodes are mainly SMEs). In the future, VEs and CNs will tend to be formed and to exist for short periods of time, i.e., the time needed to complete a business opportunity. How to structure the information for purposes of supporting the activities of temporary collaborative networks will therefore be a major difficulty in the establishment of the semantic agreements that will be the cornerstone for sharing information and knowledge.

In this paper we propose a method for the management of the knowledge structures required in the formation, operation and dissolution of a VE or CN. More specifically it focuses on the composition and segmentation of domain ontologies supporting information and knowledge management in VEs or CNs. The research focus lies essentially in the dissolution phase, which has not raised as much interest as e.g., the formation phase. It is very important that the knowledge about information structures and its use during a VE or CN operation can be returned to the VBE and reused when needed in a new VE or CN.

2   PROCESS DESCRIPTION

Assuming that a VBE possesses an Ontology Library System (OLS), during the formation process of a new VE, knowledge that is both relevant and required for supporting the VE's operations is obtained by searching and selecting data sources from this OLS. This is followed by the construction of a global domain ontology using the ontologies retrieved from the OLS (ontology composition). There will come a day when the VE will end, e.g., due to the temporary nature of the business or resulting from market flutuations. The VE will enter then in the dissolution phase. During this dissolution phase it is important to guarantee the storage of new knowledge that has been acquired during the VE's operation so that it may be reused in future (ontology decomposition).

Assuming that ontology decomposition generates meta-data, another task, which is also a goal of this research work, consists in the definition of attributes that allow for the classification of the ontologies within the library, thus supporting the effective search and selection of ontologies that occurs during the formation of news VEs. The systematization of this process is intended to contribute not only to the cyclic, dynamic and agile enrichment of a VBE's library, but also to the competences generated within each VE (see Figure 1).
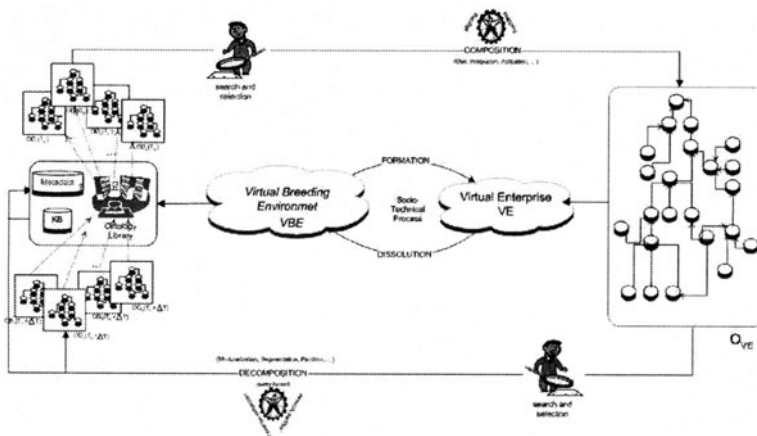

Figure 1 – Outline of Ontology Management System within VBE

The process of ontology management utilized by the OLS of the VBE, involves 5 basic steps:

SEARCH (Search the ontologies in the VBE): The goal is to find some especially relevant ontology for reuse. The search can be made by keyword, metadata, structure-based queries, supported by various evaluation schemes, or any combination of these.

SELECTION (Select the relevant ontologies): The goal is to produce a ranking of domain (concept) ontologies and respective hierarchy based on a set of metrics adopted for ontology evaluation.

COMPOSITION (Compose the ontology of the VE): The goal is to construct a global ontology that will support the VE's operation. It is thus necessary to decide on the use of an integration technique (union, inclusion, mapping, aligning, etc) that will allow for the construction of a global ontology based on the various ontologies

or parts of ontologies selected in the previous step. The resulting global ontology will therefore be more representative of the domain of interest.

DECOMPOSITION (Decompose the ontology of the VE): The goal is to break the global domain ontology resulting from the VE's constitution, into several ontologies of complementary domains. This will facilitate the reuse of such knowledge by other EVs. In this case, it is necessary to decide which are the more adequate techniques to be used in the decomposition of the global ontology, so as to enable the extraction of the reused ontologies updated and another ontology segments from others complementary domains of interest.

CLASSIFICATION (Classify and store the ontologies in the VBE): The goal is to classify the ontologies of complementary domains resulting from the segmentation of the previous step and allowing for their adequate storage in the library. Such a classification facilitates the ensuing search that is required when a new VE emerges and consequently a new global domain ontology is created. This task consists in the update of the meta-ontology.

3   ONTOLOGY LIBRARY SYSTEM

According the general orientation given for the specification of an OLS (Ding and Fensel, 2001; Noy and Musen, 2004; Noy, Rubin et al., 2004) and based on the expected capabilities of the approach taken by this work, we can say that the OLS must be supported essentially by the following infrastructural ontological components (Maedche, Motik et al., 2003; Korpilahti, 2004):

Registry – must provide a means for the search and localization of the ontologies in the VBE.

o   Search: mechanisms that support the searching and browsing of the ontologies of the OLS, namely of the Meta-Ontology (MO) are required.

o   Selection: mechanisms that support the evaluation (based on concepts and their relations), verification and consequent ranking of the ontologies found in the OLS (metrics of the evaluation) are necessary.

o   Identification: mechanisms that guarantee the (unique) identification of each domain ontology is essential.

Reuse – must provide a means for the reuse of the ontologies of the VBE.

o   Composition: mechanisms that support the automatic aligning (inclusion or integration) of the several ontology segments that are considered relevant, is required, i.e., editing of the global domain ontology with the aim of supporting the VEs operations.

Evolution – must provide methods that support an adequate and consistent evolution of the ontologies of the VBE.

o   Maintenance: mechanisms that support the addition of new ontology segments and the creation and updating of complementary domain ontologies and respective MO in the OLS are required.

o   Decomposition: mechanisms that support the extraction of the adequate ontology segments from the global ontology resulting from each VE are necessary.

o   Storage: mechanisms that support the adequate classification and registration of the domain ontology segments to be added to OLS are required.

o   Versioning: is a necessary mechanism that allows for the storage and mana-
    gement of all versions controlled copies of the ontology segments in the OLS
    (identification of the several versions of the same ontology segment and the
    information about the relevant updates between each version).

The infrastructure is based on the utilization of an ontology server whose task
consists in storing and maintaining the ontologies (as well as any of its associated
meta-information) of each VE during its operation, and in storing and maintaining
the ontology segments identified and extracted during the extinction of a VE.

As an example suppose that from the formation and operation of the VE1, the
development centre generates the OVE1 (Ontology of the VE). Once VE1's operations
have ceased, two new segments SDO1 and SDO2 are extracted. At this point VE1
contacts the OLS Ontology Registry System (ORS) and stores segments SDO1 and
SDO2. In addition to the segments themselves all their relevant meta-data, described
by the MO, is also recorded. The ORS only stores the meta-data about the extracted
segments. The segments themselves are stored elsewhere in the ontology server.
Note that, after the dissolution of VE1, OEV1 is deleted from the ontology server of
the VBE.

When a new VE is formed, its development centre now referred to as VE2, can
search for the adequate ontology segments it requires. Such segments will allow for
the composition of the global ontology OVE2 that is necessary to support the
operation of this most recent partnership. The search is based on the ontologies
meta-data.

In order to reuse SDO1 for example, the VE2 uses the VBE's ontology compo-
sition facilities, making the necessary definitions available to OVE2. Once its
operations terminate, VE2 enters the dissolution phase. In order to preserve the
knowledge acquired during its operation and provide future reuse of such knowledge,
VE2 proceeds with the extraction of the segments judged adequate for such reuse,
namely, the extraction of the original segments that have been updated but also
any additional ones considered of interest to the domain. Once segmentation is
completed, these ontology segments' meta-data is recorded in the ORS and the
ontology segments themselves are recorded elsewhere in the ontology server.

This process is repeated during the life-cycle of each new VE. With the creation
of new VEs and the consequent generation and refinement of the VBEs ontologies,
it is probable that their will be an increase in the number and reuse of ontology
segments. Thus, the process of the decomposition will pass to have several phases,
i.e., one for each ontology segment reused in the development of the global
ontology, more the necessaries to find new segments of interest in complementary
domains.

4   ONTOLOGY REGISTRY

In the ORS, the meta-data (based on the MO) will contain descriptions of all onto-
logy segments stored in the ontology sever. This description consists of information
required for the later selection of ontology segments and hence allow for the
adequate reuse and integration of these segments during the development of new
ontologies. The language used to specify either the ontology segments or the metadata
is OWL (Web ontology Language), which is a W3C recommendation.

4.1 Ranking

The ORS search is based on the identification of pertinent concepts, i.e., relevant terms in the ontology are identified by executing a keyword based search according to a set of conditions (identify concepts by class name, label or definition).

Inspired in works of Buitelaar, Eigner et al. (2004); Alani e Brewster (2005); Alani e Brewster, (2006) and also Alani, Brewster et al. (2006), we define as essentials the following metrics that would be considered in ranking and consequent selection of the ontologies of the OLS, to be applied to ontology segments resulting from the search made to OLS, with base on terms of the interest domain to the VE in formation (involving the metrics: "Covering" (MCB), "Completeness" (MCP), "Representation" (MRP) and "Proximity (MPR)), and with base on structure intended (involving the metrics: "Relationship" (MRL) and "Usage" (MUT)). These last ones have as objective complement the results of Partial Evaluation (AP), supporting a more consistent decision.

Due to space limitations, the metrics above will not be formerly presented. The interested reader is referred to the original article for more information (Alani, Brewster et al., 2006). For completeness sake however, these metrics are briefly described:

Covering (MCB) – This metric quantifies the number of classes of interest in the ontology segment. It is measured by the count of classes that are associated with a given set of labels of those classes which are identified either by exactly or partially matching the searched terms with these labels. A label of a class is considered partially matched with the searched terms if its text includes the searched term.

Completeness (MCP) – This metric measures, for a given set of terms, the degree of detail with which they are represented in the ontology segment. It is measured by the number of relations that the classes of interest (matched) possess. This count is based on the assumption that ontology segments that provide a more complete description of the concepts have greater number of relations. These relations include super-classes, sub-classes, siblings and any other declared or inferred object relation).

Representation (MRP) – This metric quantifies the importance that a concept has in a given ontology segment. It measures the centrality of classes in the ontology segment based on the presumption that if a class has a higher value of *betweenness* in the ontology segment, then this class is central to the domain represented by the ontology.

Proximity (MPR) – This metric measures the proximity between classes related to the terms of interest. Proximity is calculated according to the distance between classes. Such a distance is based on the number of relations (hierarchical and object relations) that separate any two classes. It is based on the assumption that concepts that are far apart from each other have less probability representing the domain of interest in a coherent and detailed manner.

Partial Evaluation (AP) – This is the weighed aggregation of the four metrics (MCB, MCP, MRP e MPR) described above. These metrics are calculated for each ontology segments taking into account the factors that are used to determine their relative importance in ranking the ontology segments.

Definition 1: Let M = {M[1], M[2], M[3], M[4]} = {MCB, MCP, MRP, MPR}, w*i* be the weight given to each metric and O the set of ontologies to be ranked. The values measured are normalised to a range [0..1], dividing them by the maximum value of the metrics measured for all ontologies (M[j]).

$$AP(o \in O) = \sum_{i=1}^{4} w_i \ \frac{M[i]}{\max_{1 \le j \le |O|} M[j]}$$

In addition to the metrics just described, two new metrics (MRL and MUT) are also included. They are formalised as follows:

Relationship (MRL) – This metric quantifies the extent to which one ontology is linked to others. It is measured by counting the number of ontology segments that are imported by and the number of ontology segments that import the ontology segment that is being evaluated.

Definition 2: Let I = {ontologyImported[o]} e QI = {ontologyThatImport[o]} Thus MRL(O) is the Relationship Measure of the ontology O. φ and λ are the weights given to the ontology segments imported by and that import O, respectively.

MRL(O) = φ I(O) + λQI(O)

Usage (MUT) – This metric quantifies the usage that is given to an ontology segment. It is measured by the number of individuals (instances) that are classified as members of the classes of interest. Such concepts are identified by label's exact or partial match with the search terms.

Definition 3: Let S = {Instances[C]} be the set of instances of the class C and n = E(O,T) + P(O,T), where E(O,T) and P(O,T) are the set of classes of the ontology segment O whose labels have respectively been exactly and partially matched with the searched terms (T). Thus, MUT(O) is the Usage Measure of the ontology segment O. λ and δ are the weights given.

$$MUT(O) = \frac{1}{n}\left(\lambda \sum_{i=1}^{E(o,T)} Si + \delta \sum_{j=1}^{P(o,T)} Sj \right)$$

Because it was not possible access the AKTiveRank system (Alani et al., 2006), a Java based system was developed in order to produce the results presented here. The output is generated in the CSV (Comma-Separated Value) format, which can then be easily opened and further processed by many spread sheet applications. In addition to the standard AKTiveRank metrics the MRL(O) and MUT(O) were also implemented. However, these metrics are considered as complementary to the AP result and are analyzed separately during the decision-making in choice of the ontology segment more relevant for particular context of VE in formation.

4.2   Classification

In order to facilitate the advanced query and retrieval of ontology segments, a notation management system will provide a means for domain experts to associate additional data to the ontology segment. Such information, provided in the form of meta-data, will allow for the extensive documentation of the ontology itself. It will include data on such things as: provenance, how the ontology was constructed, recommendations of how to proceed with its extensions, which nomination policy was followed, what are the organizational principles and functions, etc.
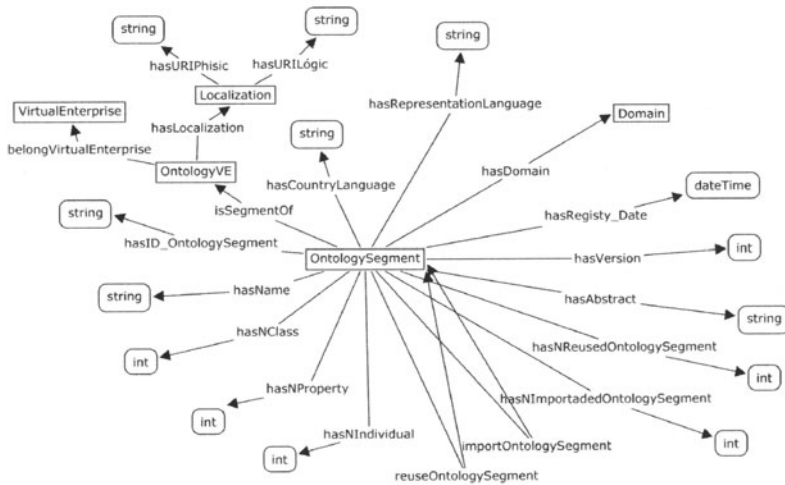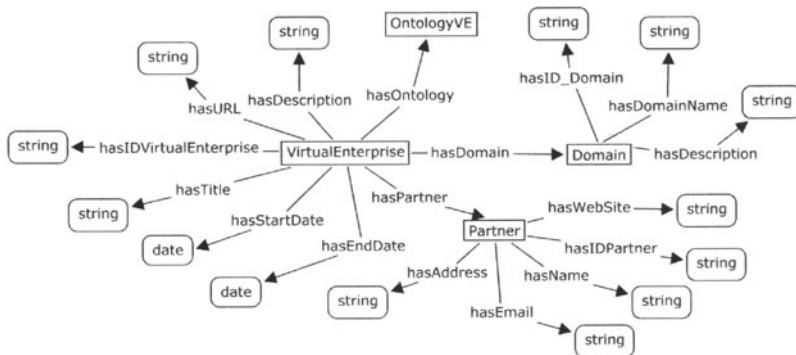
Figure 2 – Conceptual Map of Meta-Ontology

Figure 3 – Conceptual Map of Domain Ontology

In order to facilitate the adequate classification of ontology segments when they are registry in the ORS, the ontology segments will be annotated using the concepts in the MO (presented in form of conceptual map in Figure 2). An additional Domain Ontology (see Figure 3) was also created in order to allow for a more uniform annotation of the ontology segments. Both the MO and Domain Ontology were developed in OWL format.

5   TESTS

In order to test the ranking of the ontologies in the OLS using of the metrics described in section 4.1, two small ontologies that share a percentage of common concepts (labels), but with a relatively different hierarchy structure and set of properties (object property) were used. The ontologies are identified as "community_v2" (O1) and "knowledge_community_v2" (O2). Ontology O2 imports an additional ontology designated as "virtual_network". In both ontologies several instances (individuals) were added. All ontologies were created in the OWL language using the Protégé editor. During the development of these ontologies each class was given a generic designation using a numeric based denomination (C1, C2, etc). In addition

to this each class was also assigned a label that identified it (see Figure 4 and Figure 5). This use of labels was done in order to allow for experimentation with the same hierarchy structure but with different sets of labels.
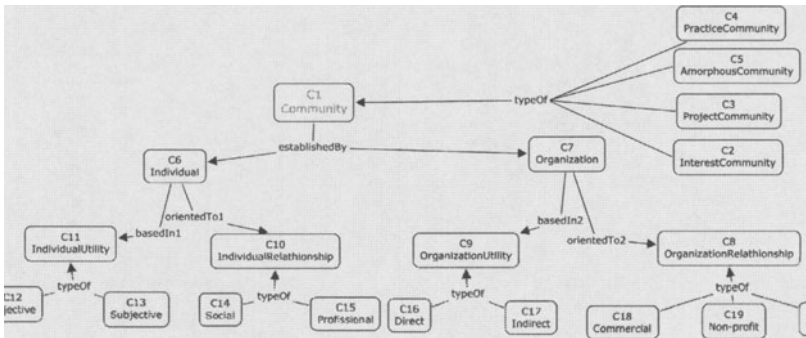

Figure 4 – Conceptual Map of "community_v2.owl" Ontology
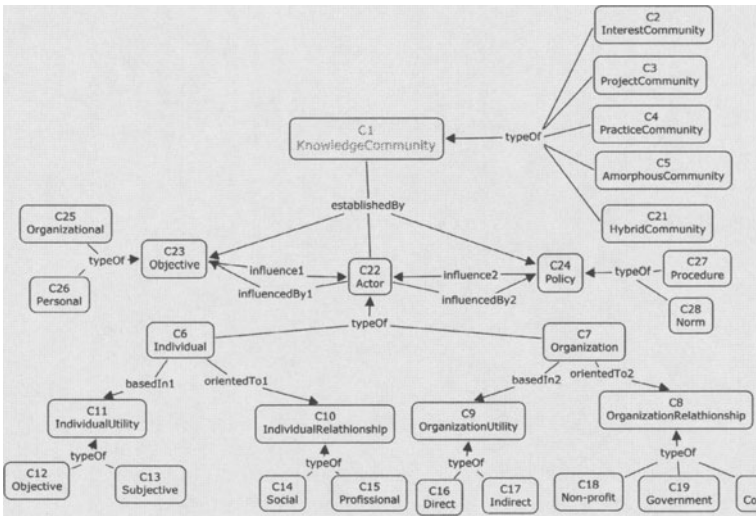

Figure 5 – Conceptual Map of "knowledge_community_v2.owl" Ontology

Table 1 shows the results obtained when calculating the metrics described in 4.1, each of which pertains to a different set of search terms (T). The results of the AP metric were obtained using the following weights: $w_{MCB}=w_{MCP}=0.3$ e $w_{MRP}=w_{MPR}=0.2$ such that $w_{MCB}+w_{MCP}+w_{MRP}+w_{MPR}=1$. The metrics that are considered more relevant to the selection (and therefore ranking) of the ontology are the MCB and MCP metrics. This justifies the choice of assigning larger values to the weights of the MCB and MCP metrics than those of the MRP and MPR metrics. In the case of the MCB, we considered the weights given to the exact matches greater than those given to partial matches (respectively 0.6 to 0.4). Relatively to the MCP, all four parameters (respectively super-classes, sub-classes, relations and siblings) were given equitable weight.

The results show that for all sets of searched terms ontology O2 is always ranked above ontology O1. In other words ontology O2 is considered to be of greater

relevance than ontology O1 for this specific search. These results are as expected. Observer that ontology O2 is denser and more complete than ontology O1. It also has a higher frequency of use of the term "community". Because all searches include this term it is natural that O2 be ranked above O1. Note also that even when the set of search terms includes other equally relevant terms ("individual", "organization" and "social") the ranking does not change significantly.

Table 1 – Application of metrics to ontologies O1 and O2

T={community}							
source	MCB	MCP	MRP	MPR	MRL	MUT	AP
./metrics2/community_v2.owl	2.20	1.25000	0.00000	0.40000	0.00	0.80	0.35357
./metrics2/knowledge_community_v2.owl	2.80	2.42857	1.42857	0.90476	0.50	0.34	0.69116
T={community,INDIVIDUAL}							
source	MCB	MCP	MRP	MPR	MRL	MUT	AP
./metrics2/community_v2.owl	3.60	0.96875	0.62500	0.57143	0.00	0.85	0.37513
./metrics2/knowledge_community_v2.owl	4.20	1.95000	1.46667	0.64815	0.50	0.52	0.52848
T={community,INDIVIDUAL,orGaniZatioN}							
source	MCB	MCP	MRP	MPR	MRL	MUT	AP
./metrics2/community_v2.owl	5.00	0.86364	0.90909	0.50909	0.00	0.87	0.30107
./metrics2/knowledge_community_v2.owl	6.80	1.50000	3.09688	0.51264	0.50	0.61	0.49582
T={community,INDIVIDUAL,orGaniZatioN, social}							
source	MCB	MCP	MRP	MPR	MRL	MUT	AP
./metrics2/community_v2.owl	5.60	0.83333	1.15278	0.47222	0.00	0.85	0.30905
./metrics2/knowledge_community_v2.owl	7.40	1.44118	3.29202	0.48206	0.50	0.61	0.48544

In the case of MRL metric both parameters were given an equal weight of 0.5. In respect to the MUT metric greater importance was given to instances belonging to classes that were identified by exact matches than those by partial matches. The value assigned to the metrics of exact and partial matches is respectively 0.6 and 0.4. It is important to reiterate that these metrics serve only as a complementary means of better assessing the ranking of the ontologies provided by do the MCB, MCP, MRP and MPR metrics. The MUT metric will be useful, for example, in case the AP scores between two or more ontologies are either tied or very close. In such cases we would opt for the ontology that demonstrates a greater degree of usage (large values of MUT). The metric MRL, in its turn, will help with the selection, development and hence adaptation of the ontology segments. During this development process larger value of MRL are an indication that the selected ontologies are based on a greater number of reusable (and arguably more stable) ontologies are therefore better candidates for reuse themselves.

6   DISCUSSION AND FUTURE WORK

This paper presented an approach used in the management of multiples ontologies within the context of a VBE. A meta-Ontology and Domain-Ontology were also presented which support the annotation; storage and retrieval of ontologies in the OLS were also presented. A set of metrics for the ranking and selection of ontology segments were also described. These metrics were exemplified by calculating their values with respect to two ontologies in an experiment. The resulting ranking of this experiment was discussed.

In addition to ontology evaluation and ranking, a segmentation algorithm is currently being developed. This algorithm will automate the decomposition and

extraction of ontology segments from a global ontology during the dissolution of a VE. These extracted segments allow for the collection, comparison, recording and version control of previously reusable ontology segments. In addition to this it will also allow for the identification and registration of new ontology segments (according to the domains of interest as is specified in Domain Ontology) from ontologies that resulted from the compositions and consequent evolution of a VE's operation.

7   REFERENCES

1.   Afsarmanesh, H. e L. M. Camarinha-Matos (2005). A framework for management of virtual organization breeding environments. Collaborative Networks and their Breeding Environments, (PRO-VE'05), Valencia, Spain, Springer.
2.   Alani, H. e C. Brewster (2005). Ontology Ranking based on the Analysis of Concept Structures. Third International Conference on Knowledge Capture (K-Cap), Banff, Alberta, Canada.
3.   Alani, H. e C. Brewster (2006). Metrics for Ranking Ontologies. WWW2006, Edinburgh, UK.
4.   Alani, H., C. Brewster, et al. (2006). Ranking Ontologies with AKTiveRank. 5th International Semantic Web Conference (ISWC), Georgia, USA.
5.   Bacquet, J., P. Fatelnig, et al. (2004). An outlook of future research needs on networked organizations. Virtual Enterprises and Collaborative Networks (PROVE'O4), Toulouse, France, Kluwer Academic Publishers.
6.   Buitelaar, P., T. Eigner, et al. (2004). OntoSelect: A Dynamic Ontology Library with Support for Ontology Selection. International Semantic Web Conference, Hiroshima, Japan.
7.   Ding, Y. e D. Fensel (2001). "Ontology Library Systems: The key to successful Ontology Re-use." 93 - 112.
8.   Korpilahti, T. (2004). Arquitecture for distributed development of an ontology library. Department of Computer Science and Engineering. Helsinki, Helsinki University of Technology. Master of Science in Technology: 60.
9.   Maedche, A., B. Motik, et al. (2003). An Infrastructure for Searching, Reusing and Evolving Distributed Ontologies. WWW2003, Budapest, Hungary, ACM.
10. Noy, N. F. e M. A. Musen (2004). "Ontology versioning in an ontology management framework." IEEE Intelligent Systems 19(4): 6.
11. Noy, N. F., D. Rubin, et al. (2004). "Making Biomedical Ontologies and Ontology Libraries Work." EEE Intelligent Systems 19(6): 78 - 81.
12. Porter, C. E. (2004). "A Typology of Virtual Communities: A Multi-Disciplinary Foundation for Future Research." Journal of Computer-Mediated Communication (JCMC) 10((1), Article 3).