The HCONE Approach to Ontology Merging

Konstantinos Kotis[1], George A. Vouros[1]

[1]Dept. of Information & Communications Systems Engineering,
University of the Aegean,
Karlovassi, Samos,
83100, Greece
{kkot, georgev}@aegean.gr

Abstract. Existing efforts on ontology mapping, alignment and merging vary from methodological and theoretical frameworks, to methods and tools that support the semi-automatic coordination of ontologies. However, only latest research efforts "touch" on the *mapping /merging* of ontologies using the whole breadth of available knowledge. This paper aims to thoroughly describe the HCONE approach on ontology merging. The approach described is based on (a) capturing the intended informal interpretations of concepts by mapping them to WordNet senses using lexical semantic indexing, and (b) exploiting the formal semantics of concepts' definitions by means of description logics' reasoning services.

1 Introduction

Ontologies have been realized as the key technology to shaping and exploiting information for the effective management of knowledge and for the evolution of the Semantic Web and its applications. In such a distributed setting, ontologies establish a common vocabulary for community members to interlink, combine, and communicate knowledge shaped through practice and interaction, binding the knowledge processes of creating, importing, capturing, retrieving, and using knowledge. However, it seems that there will always be more than one ontology even for the same domain. In such a setting where different conceptualizations of the same domain exist, information services must effectively answer queries bridging the gaps between their formal ontologies and users' own conceptualizations. Towards this target, networks of semantically related information must be created at-request. Therefore, coordination (i.e. mapping, alignment, merging) of ontologies is a major challenge for bridging the gaps between agents (software and human) with different conceptualizations.

In [1] an extensive discussion about the way "semantics" are introduced, formalized and exploited in the semantic web, shows that coordination of ontologies using semantic knowledge can be achieved through several methods, depending on *where the semantics are* across the *semantic continuum*: From humans' minds, to their explicit but informal description, their formal description intended for human use, and finally, to their explicit and formal specification intended for machine utilization. The further we move along the continuum, from implicit to formal, explicit semantics, ambiguity is reduced and automated inference is made possible, regarding fully automated semantic interoperation and integration. Looking for methods that will fully automate the mapping, alignment and merging processes between ontologies, today we devise methods that are located in the middle of this continuum.

There are many works devoted to coordinating ontologies that exploit linguistic, structural, domain knowledge and matching heuristics. Recent approaches aim to exploit all these types of knowledge and further capture the intended meanings of terms by means of heuristic rules [2].

The HCONE [3] approach to merging ontologies exploits all the above-mentioned types of knowledge. In a greater extent than existing approaches to coordinating ontologies, this approach gives much emphasis on "uncovering" the intended informal interpretations of concepts specified in an ontology. Linguistic and structural knowledge about ontologies are exploited by the Latent Semantics Indexing method (LSI) for associating concepts to their informal, human-oriented intended interpretations realized by WordNet senses. Using concepts' intended semantics, the proposed method translates formal concept definitions to a common vocabulary and exploits the translated definitions by means of description logics' reasoning services. The goal is to validate the mapping between ontologies and find a minimum set of axioms for the merged ontology.

Our choice of description logics is motivated by the need to find the minimum set of axioms needed for merging, and to test the formal consistency of concepts' definitions by means of classification and subsumption reasoning services.

According to the suggested approach and with respect to the semantic continuum, humans are involved in the merging process in two stages: In capturing the intended semantics of terms by means of informal definitions (supported by LSI), and in clarifying relations between concepts in case such relations are not stated formally.

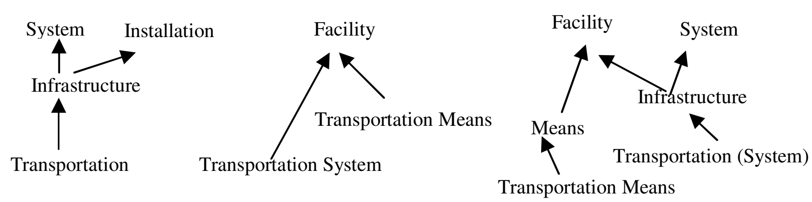
The paper is structured as follows: Section 2 formalizes the problem of semantically merging ontologies. Section 3 describes the HCONE approach to merging ontologies. Section 4 discusses the proposed approach, with remarks, insights on the relation of the proposed approach to other approaches, and future work.

2 The Problem Specification

In order to have a common reference to other approaches, we formulate the problem by means of definitions and terms used in [2].

An ontology is considered to be a pair O=(S, A), where S is the ontological signature describing the vocabulary (i.e. the terms that lexicalize concepts and relations between concepts) and A is a set of ontological axioms, restricting the intended interpretations of the terms included in the signature. In other words, A includes the formal definitions of concepts and relations that are lexicalized by natural language terms in S. This is a slight variation of the definition given in [2], where S is also equipped with a partial order based on the inclusion relation between concepts. In our definition, conforming to description logics' terminological axioms, inclusion relations are ontological axioms included in A. It must be noticed that in this paper we deal with inclusion and equivalence relations among concepts.

Ontology mapping from ontology $O_1 = (S_1, A_1)$ to $O_2 = (S_2, A_2)$ is considered to be a morphism $f{:}S_1 \rightarrow S_2$ of ontological signatures such that $A_2 \models f(A_1)$, i.e. all interpretations that satisfy O_2's axioms also satisfy O_1's translated axioms. Consider for instance the ontologies depicted in Figure 1.



O_1= ({System, Infrastructure, Installation,Transportation},
{Transportation $\sqsubseteq$ Infrastructure, Infrastructure $\sqsubseteq$ Installation, Infrastructure $\sqsubseteq$ System})

O_2= ({Facility, Transportation System, Transportation Means, exploit},
{Transportation System $\sqsubseteq$ Facility, Transportation Means $\sqsubseteq$ Facility $\sqsubseteq$ exploit.TransportationSystem })

O_3= ({System, facility, Means, Installation, Infrastructure, Transportation System, Transportation, Transportation Means, exploit},
{Transportation $\equiv$ Transportation System, Facility $\equiv$ Installation, Infrastructure $\sqsubseteq$ System $\sqcap$ Facility,
Transportation System $\sqsubseteq$ Infrastructure $\sqcap$ Facility,
Transportation Means $\sqsubseteq$ Means $\sqcap$ exploit.TransportationSystem , Means $\sqsubseteq$ Facility})

Fig. 1. Example Ontologies

Given the morphism f such that *f(Infrastructure)=Facility* and *f(Transportation)=Transportation System*, it is true that $A_2 \models \{f(Transportation) \sqsubseteq f(Infrastructure)\}$, therefore f is a mapping. Given the morphism f', such that *f'(Infrastructure)=Transportation System* and *f'(Transportation)= Transportation Means*, it is not true that $A_2 \models \{f(Transportation) \sqsubseteq f(Infrastructure)\}$, therefore f' is not a mapping.

However, instead of a function, we may articulate a set of binary relations between the ontological signatures. Such relations can be the inclusion ($\sqsubseteq$) and the equivalence ($\equiv$) relation. For instance, given the ontologies in Figure 1, we can say that *Transportation≡Transportation System, Installation≡Facility* and *Infrastructure ⊑ Facility*. Then we have indicated an alignment of the two ontologies and we can merge them. Based on the alignment, the merged ontology will be ontology O_3 in Figure 1. It holds that $A_3 \models A_2$ and $A_3 \models A_1$.

Looking at Figure 1 in an other way, we can consider O_3 to be part of a larger intermediary ontology and define the alignment of ontologies O_1 and O_2 by means of morphisms $f_1 : S_1 \rightarrow S_3$ and $f_2 : S_2 \rightarrow S_3$. Then, the merging of the two ontologies [2] is the minimal union of ontological vocabularies and axioms with respect to the intermediate ontology where ontologies have been mapped.

Therefore, the ontologies merging problem (OMP) can be stated as follows: *Given two ontologies find an alignment between these two ontologies, and then, get the minimal union of their (translated) vocabularies and axioms with respect to their alignment.*

3 The HCONE Method to Solving the OMP

As it is shown in Figure 2, WordNet plays the role of an "intermediate" in order a morphism to be found. We consider that each sense in a WordNet synset describes a concept. WordNet senses are related among themselves via the inclusion (hyponym – hyperonym) relation. Moreover, terms that lexicalize the same concept (sense) are considered to be equivalent through the synonym relation.

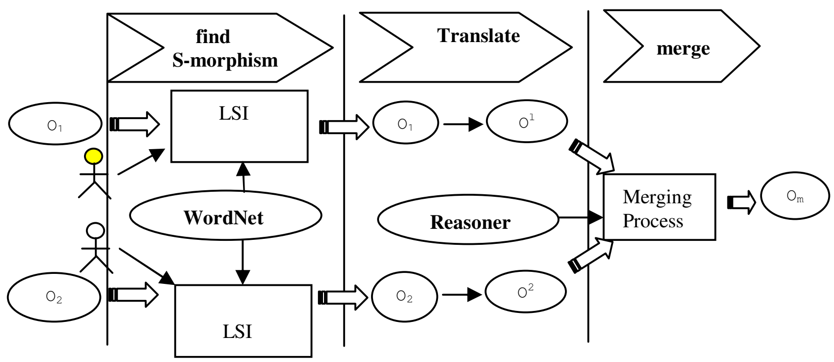


Fig. 2. The HCONE approach towards the OMP

It must be noticed that we do not consider WordNet to include any intermediate ontology, as this would be very restrictive for the specification of the original ontologies (i.e. the method would work only for those ontologies that preserve the inclusion relations among WordNet senses).

Therefore, we consider that there is an intermediate ontology "somewhere there" including a vocabulary with the lexicalizations of the specific senses of WordNet synsets we are interested on, and axioms that respect the set of axioms of the original ontologies. We will call this ontology *hidden intermediate*. It is important to notice that only part of this ontology will be uncovered through concept mappings: actually, the part that is needed for merging the source ontologies.

To find the mapping from each ontology to the hidden intermediate, we use a morphism (we call it *s-morphism*, symbolized by f_s), which is based on the lexical semantic indexing (LSI) method. Using the LSI method, each ontology concept is associated with a set of graded WordNet senses. For instance, the concept "facility" is associated with the five senses that WordNet assigns to the term "facility", whose meaning range from "something created to provide a service" to "a room equipped with washing and toilet facilities". The highest graded sense expresses the most possible informal meaning of the corresponding concept. This sense expresses the intended interpretation of the concept specification and can be further validated by a human. In case a human indicates an association to be the most preferable, then this sense is considered to capture the informal intended meaning of the formal ontology concept. Otherwise, the method considers the highest graded sense as the concept's intended interpretation. Given all the preferred associations from concepts to WordNet senses, we have captured the intended interpretation of ontology concepts.

Using the intended meanings of the formal concepts, we construct an ontology $O^n=(S^n, A^n)$, $n=1,2,$ where, S^n includes the lexicalizations of the senses associated to the concepts[1] of the ontology $O_n=(S_n, A_n)$, $n=1,2,$ and A^n contain the translated inclusion and equivalence relations between the corresponding concepts. Then, it holds that $A^n \models f_s(A_n)$ and the ontology $O^n=(S^n, A^n)$ with the corresponding associations from O_n *to* O^n, is a model of $O_n=(S_n, A_n)$, $n=1,2\ldots$. These associations define a mapping from O_n to O^n.

Having found the mappings with the hidden intermediate ontology, the translated ontologies can be merged, taking into account the axioms A^1 and A^2 (which are the translated axioms of A_1 and A_2). The merging decisions are summarized in Table 1.

Table 1. HCONE-Merge Algorithm table summary

Concept & Role **Names**[2]	Concept Mapping to WordNet **Senses**[3]	**Action**
Match	No match	Rename concepts
Match	Match	Merge concept definitions
No match	Match	Merge concept definitions in a single concept named by the term lexicalizing their corresponding WordNet sense
No match	No match	Classify Concepts

3.1 Mapping and Merging through the Semantic Morphism (s-morphism)

To find the mapping from an ontology to the hidden intermediate, we use the semantic morphism (*s-morphism*, symbolized by *fs*), which, as already pointed, is based on the lexical semantic indexing (LSI) method.

LSI [5] is a vector space technique for information retrieval and indexing. It assumes that there is an underlying latent semantic structure that it estimates using statistical techniques. It takes a large matrix of term-document association data and constructs a semantic space. In our case the nXm space comprises the n more frequently occurred terms of the m WordNet senses the algorithm focuses on (later on we explain which senses constitute the focus of the algorithm). Lexical Semantic Analysis (LSA) allows the arrangement of the semantic space to reflect the major associative patterns in the data. As a result, terms that did not actually appear in a sense may still end up close to the sense, if this is consistent with the major patterns of association in the data [5]. Position in the space then serves as the new kind of semantic indexing. Therefore, it must be emphasized that although LSI exploits structural information of ontologies and WordNet, it ends up with semantic associations between terms.

Given an ontology concept, retrieval aims to locate a point in space that is close to the sense that expresses the intended meaning of this concept. The query to the retrieval mechanism is constructed by the concept names of all concepts in the vicinity of the given concept.

[1] Future work concerns mapping domain relations to WordNet senses as well.
[2] Match in this case means linguistic match of the concept names from the two ontologies.
[3] Match means that both concepts have been mapped to the same WordNet sense

To support this process, as already explained, we exploit the WordNet lexical database to match formal descriptions of concepts with word senses in WordNet. Using the lexicalizations of these senses, the ontology is translated to the hidden intermediate ontology. The steps of the algorithm for finding the semantic morphism are the following:

1. Choose a concept from the ontology. Let C be the concept name.
2. Get all WordNet senses $S_1, S_2, \ldots, S_m$, lexicalized by C', where C' is a linguistic variation of C. These senses provide the *focus of the algorithm for C*.
3. Get the hyperonyms' and hyponyms' of all C' senses.
4. Build the "*semantic space*": An nXm matrix that comprises the n more frequently occurred terms in the *vicinity* of the m WordNet senses found in step 2.
5. Build a query string using the terms in the *vicinity* of C.
6. Find the ranked associations between C and C' senses by running the Latent Semantics Analysis (LSA) function and consider the association with the highest grade. LSA uses the query terms for constructing the query string and computes a point in the semantic space constructed in step (4).

This algorithm is based on assumptions that influence the associations produced:

– Currently, concept names lemmatization and morphological analysis is not sophisticated. This implies that in case the algorithm does not find a lexical entry that matches a slight variation of the given concept name, then the user is being asked to provide a synonym term. However, in another line of research we produce methods for matching concept names based on a 'core set' of characters [4].

– Most compound terms have no senses in WordNet, thus we can only achieve an association for each component of the term (which is a partial indication of the intended meaning of the whole term). Currently, we consider that the compound term lexicalizes a concept that is related (via a generic relation *Relation*) to concepts that correspond to the single terms comprising the compound term. For instance, the concept lexicalized by "Transportation Means" is considered to be related to the concepts lexicalized by "Transportation" and "Means". It is assumed that humans shall clarify the type of relations that hold between concepts. Such a relation can be the inclusion, equivalence or any other domain relation. In general, in case a compound term C cannot be associated with a sense that expresses its exact meaning, then the term C is associated with concepts H_n, n=1,2… corresponding to the single words comprising it. Then, C is considered to be mapped in a virtual concept C_w of the intermediate ontology, while H_n are considered to be included in the ontological signature of the intermediate ontology and the axiom $C_w \sqsubseteq \sqcap_{Hn} \forall Relation.H_n$, n=1,2… is considered to be included in ontological axioms of the intermediate ontology. For instance, *"Transportation Means"* is considered to correspond to a virtual concept of the intermediate ontology, while the axiom $TransportationMeans \sqsubseteq Relation.Transportation \sqcap Relation.Means$ is considered to be an axiom of this ontology. Given that *Means* subsumes *TransportationMeans,* and the *Relation* to *Trasportation* is *function,* the axiom becomes *Transportation-Means* $\sqsubseteq$ *Means* $\sqcap$ *function.Transportation.*

This treatment of compound terms is motivated by the need to reduce the problem of mapping these terms to the mapping of single terms. Then we can exploit the translated formal definitions of compound terms by means of description logics reasoning services for testing equivalence and subsumption relations between concepts definitions during ontologies alignment and merging.

– The performance of the algorithm is related to assumptions concerning the information that has to be used for the computation of the (a) "semantic space", and (b) query terms.

– The implementation of LSI that we are currently using, as it is pointed by the developers[4], works correctly when the nXm matrix corresponding to the semantic space has more than 4 and less than 100 WordNet senses. This case occurs frequently, but we resolve it by extending the vicinity of senses.

The semantic space is constructed by terms in the *vicinity* of the senses S_1, S_2,…S_m that are in focus of the algorithm for a concept C. Therefore, we have to decide what constitutes the *vicinity* of a sense for the calculation of the semantic space. In an analogous way we have to decide what constitutes the *vicinity* of an ontology concept for the calculation of the query string. The goal is to compute valid associations without distracting LSI with "noise" and by specifying vicinities in an application independed way.

Table 2. Algorithm's design assumptions (The switches with the asterisk are always activated)

Semantic Space Variations	☑ concept's name*	The term C' that corresponds to C. C' is a lexical entry in WordNet
	☑ concept's senses*	Terms that appear in C' WordNet senses
	☑ hyperonyms & hyponyms	Terms that constitute hyperonyms / hyponyms of each C' sense.
	☑ hyperonyms' / hyponyms' senses	Terms that appear in hyper(hyp)onyms of C' senses
Query Terms Variation	☑ primitive parents*	Concept's C primitive parents.
	☑ taxonomy parents	Concepts that subsume C and are immediate parents of C (subsumers of C).
	☑ children*	Concepts that are immediate children of C (subsumed by C)
	☑ related concepts	Concepts that are related to C via domain specific relations
	☑ WordNet Senses	The most frequent terms in WordNet senses that have been associated with the concepts in the vicinity of C.

[4] KnownSpace Hydrogen License: This product includes software developed by the Know Space Group for use in the KnownSpace Project (http://www.knownspace.org)

Towards this goal we have ran a set of experiments by activating / deactivating the "switches" shown in Table 2, thus specifying "vicinity" based on structural features of the ontology and WordNet.We have run experiments both in small (10 concepts) and large ontologies (100 concepts) for the transportation domain. The ontology presented in this paper comprises about 10 concepts. It must be noticed that using the proposed method, small ontologies are considered to be harder to be mapped since the available information for performing the semantic analysis is limited. By the term "small" ontologies" we denote ontologies for which the vicinity of ontology concepts includes a limited number of concepts for the construction of the query string. On the contrary, in "large" ontologies, the query string can include sufficient information for computing the intented sense of ontology concepts.

Furthermore, the small ontology allowed us to control and criticize the results. To perform our experiments we have distinguished several cases whose results have been measured by method's recall and precision. These cases correspond to different WordNet senses and concepts' vicinity definitions. Table3 shows two cases that resulted to high (balanced case) and quite low precision (all-activated case). Recall in all these cases was constantly 90% due to one compound term that could only be partially associated to a WordNet sense. Similar results have been given by larger ontologies.

Table 3. A balanced combination of senses and concepts' vicinities (defined by the activated switches for the computation of the semantic space and queries, respectively) resulted to the highest precision percentage of 90%

	Variations of LSI algorithm – Design Implications		Precision
Balanced Case	Space Variations	☑ concept	90%
		☑ concept senses	
		☐ hyper(hyp)onyms	
		☐ hyper(hyp)onyms senses	
	Query Terms Variation	☑ primitive parents	
		☐ taxonomy parents	
		☑ children	
		☐ WordNet Senses	
		☑ related concepts	
All-activated Case	Space Variations	☑ concept	50%
		☑ concept senses	
		☑ hyper(hyp)onyms	
		☑ hyper(hyp)onyms senses	
	Query Terms Variation	☑ primitive parents	
		☑ taxonomy parents	
		☑ children	
		☑ WordNet Senses	
		☑ related concepts	

The balanced case corresponds to a "balanced amount" of information for computing the semantic space and for constructing the queries. The conjecture that LSI can be distracted in large semantic spaces by, what we may call, *semantic noise*, has been proved in test cases where the semantic space has been computed taking into account the WordNet senses of the hyperonyms and/or hyponyms of the senses in focus. By reducing the amount of information in the semantic space we actually achieved to get more hits. Experiments imply that to compute correct concept-senses associations, LSI must consider senses that are "close" to the meaning of the concepts in the hidden intermediate ontology, otherwise noise (or ellipsis of information) can seriously distract computations due to influences from other domains/conceptualizations. A similar case occurs when the query string includes terms that appear in the WordNet senses associated with the super-concepts and sub-concepts of an ontology concept. For this reason, the balanced case shown in Table 3 does not consider these WordNet senses. The balanced case has been specified manually, by the proper examination of experiments. In a latest work of ours, we are investigating techniques for automatically tuning the mechanism to maximize the precision.

Having found the associations between the ontology concepts and WordNet senses, the algorithm has found a semantic morphism between the original ontologies and the hidden intermediate ontology. The construction of the intermediate ontology with the minimal set of axioms results in ontologies' merging.

For instance, as it is shown in Figure 3, given the morphisms produced, it holds that:

- For ontology O_1

  $f_s(System) = System_{1,}$
  $f_s(Installation) = Facility_{1,}$
  $f_s(Infrastructure) = Infrastructure_1$, and
  $f_s(Transportation) = TransportationSystem_1$.

- For ontology O_2

  $f_s(Facility) = Facility_1$,
  $f_s(Transportation\ System) = TransportationSystem_1$, and
  $f_s(Transportation\ Means) = TransporationMeans_W$ *{virtual concept}*
  $f_s(Means) = Means_1$
  $f_s(Transportation) = Transportation_2$

The indices of the associated terms indicate the WordNet senses that provide the informal intended meanings of concepts. Notice that the intended interpretation of the concept Transportation in O_2 is different from the intended interpretation of the homonym concept in O_1.

Both ontologies are being translated using the corresponding WordNet senses' lexicalizations and are being merged. We must notice that the compound term *Transportation Means* has been related with the concept *Transportation* (the relation has been specified to be *function)* and with the concept *Means* (via the subsumption relation). This definition is in conjunction to the definition given in O_2, where Transportation Means are defined to be entities that exploit the Transportation System.

The new ontology will incorporate the mappings of the original concepts, the translated axioms of O_1 and O_2, modulo the axioms of the intermediate ontology.

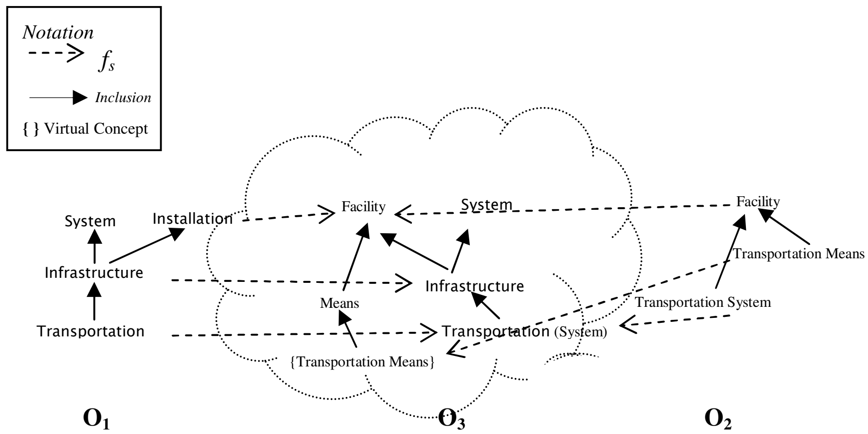


Fig. 3. S-morphism and the intermediate ontology

Therefore, the merged ontology is $O_m = (S_m, A_m)$, where:

S_m=*{System, facility, Means, Installation, Infrastructure, Transportation System, Transportation, Transportation Means, exploit},*
A_m=*{Transportation ≡TransportationSystem,*
Facility ≡Installation, Infrastructure ⊑ System ⊓ Facility,
TransportationSystem ⊑ Infrastructure Means ⊑ Facility,
TransportationMeans ⊑Means ⊓ function.Transportation-O2
⊓ exploit.TransportationSystem }

It must be noticed that the concepts *Transportation* and *Transportation System* have the same intended interpretation, and therefore are considered equivalent. According to Table 1, the merging of their formal definitions results to:

TransportationSystem ⊑ Infrastructure ⊓ Facility

However, the description logics classification mechanism considers the axiom *TransportationSystem ⊑ Facility* to be redundant. Therefore O_3 contains only the axiom *TransportationSystem ⊑ Infrastructure*. Doing so, the merged ontology contains only the minimal set of axioms resulting from original ontologies mapping.

Furthermore, according to Table 1, the concept Transportation of O_2 will be renamed to Transportation-O_2 since it corresponds to a sense that is different to the sense of the homonym concept Transportation in O_1. This latter concept, based on the morphism, has been renamed to TransportationSystem.

4 Concluding Remarks

As already explained in section 2, mapping between ontologies has a close relation to the merging of ontologies. Mapping may utilize a reference ontology but it can also be point-to-point (non mediated). In either case it must preserve the semantics of the mapped ontology. The merging process takes into account the mapping results [6] in order to resolve problems concerning name conflicts, taxonomy conflicts, etc between the merged ontologies.

To accomplish a mapping between two conceptual models, a matching algorithm is required which will eventually discover these mappings. *Matching* can be distinguished in syntactic, structural and semantic matching depending on the knowledge utilized and on the kind of the similarity relation used [7]. Syntactic matching involves the matching of ontology nodes' labels, estimating the similarity among nodes using syntactic similarity measures, as for instance in [8]. Minor name and structure variations can lead the matching result astray. On the other hand, structural matching involves matching the neighbourhoods of ontology nodes, providing evidence for the similarity of the nodes themselves. Semantic matching explores the mapping between the meanings of concept specifications exploiting domain knowledge as well. Semantic matching specifies a similarity relation in the form of a semantic relation between the intensions of concepts [9]. Semantic matching may also rely to additional information such as lexicons, thesaurus or reference ontologies incorporating semantic knowledge (mostly domain dependent) into the process.

Instance based approaches to mapping and merging ontologies, which contrast techniques for merging non-populated ontologies, exploit the set-theoretic semantics of concept definitions in order to uncover semantic relations among them. However, such approaches deal with specific (quite restricted) domains of discourse, rather than with the semantics of the statements themselves. Therefore, these approaches are useful in cases where information sources are rather stable (where the domain of discourse does not change frequently) or in cases where information is "representative" (e.g., as it is required in FCA-Merge) for the concepts specified.

There are a variety of research efforts towards coordinating ontologies. According to [10] and [11] there is not a "best tool" or method, since there is not always the case that it will fit every users' or applications' needs. To comment however on such efforts, we conjecture that several criteria could be considered such as:

a) The kind of *mapping architecture* they provide:(i) point-to-point mapping or mediated mapping, (ii) top-down or bottom up mapping, considering techniques applied to the intensions of concepts (non-populated ontologies) or to the extensions of concepts (populated ontologies), respectively.
b) The *kind of knowledge* (structural, lexical, domain) used for node matching, i.e. i) techniques that are based on the syntax of labels of nodes and to syntactic similarity measures, ii) techniques that are based on the semantic relations of concepts and to semantic similarity measures, and iii) techniques that rely on structural information about ontologies.
c) The *type of result* corresponding algorithms produce: For instance, a mapping between two ontologies or/and a merged ontology

d) Additional information sources consulted during the mapping/merging process, for instance, thesaurus, lexicons.
e) The level of user involvement: How and when the user is involved in the process.

Table 4 summarises the existing efforts to ontologies' coordination using the above issues. A careful examination of the table shows that each effort focuses on certain important issues. The HCONE method to merging ontologies, borrowing from the results of the reported efforts, focuses on all of the issues mentioned above.

In particular, we have realised that efforts conforming to mediated mapping and merging [12][13] will possibly not work, since a reference ontology (that preserves the axioms of the source ontologies) may not be always available or may be hard to be constructed (especially in the "real world" of the SemanticWeb). On the other hand, point-to-point efforts are missing the valuable knowledge (structure and domain) that a reference ontology can provide in respect to the semantic similarity relations between concepts. The proposed HCONE merging process assumes that there is a hidden intermediate reference ontology that is build on the fly using WordNet senses, expressing the intended interpretations of ontologies' concepts, and user specified semantic relations among concepts.

Although bottom-up approaches [12], [13], [14] rely on strong assumptions concerning the population of ontologies, they have a higher grade of precision in their matching techniques since instances provide a better representation of concepts' intended meaning in a domain. However, using WordNet senses we provide an informal representation of concepts' intensions (i.e. of the conditions for an entity to belong in the denotation of a concept, rather than the entities themselves).

More importantly, we have identified that apart from [9], [15] all efforts do not consult significant domain knowledge. However, to make use of such knowledge, additional information must be specified in the ontology. WordNet is a potential source of such information [9]. However, utilizing this source implies that the domain ontologies must be consistent to the semantic relations between WordNet senses, which is a very restrictive (if not prohibiting) condition to the construction of source ontologies.

HCONE exploits WordNet, which is an external (to the source ontologies) natural language information source. The proposed HCONE method consults WordNet for lexical information, exploiting also structural information between senses in order to obtain interpretations of concepts (i.e. the informal human oriented semantics of defined terms). Other efforts such as [8], [14], [16] have used additional information sources but only [9] have used WordNet for lexical and domain knowledge.

A complete automated merging tool is not the aim of this research. Since we conjecture that in real environments such as the Semantic Web humans' intended interpretations of concepts must always be captured, the question is where to place this involvement. Existing efforts [12][15][14], place this involvement after the mapping between sources ontologies has been produced, as well as during, or at the end of the merging method. The user is usually asked to decide upon merging strategies or to guide the process in case of inconsistency. Some other efforts head towards automatic mapping techniques [9], [8], [13] but they have not shown that a consistent and automatic merging will follow.

Table 4. Issues concerning existing ontology mapping/merging tools

	Mapping Architec-ture	Kind of knowledge used	Type of result	N.L Informa-tion	User Involve-ment
ONIONS [12]	Mediated Bottom-up	Syntactic	Mapping & Merging	No	Semi-automatic
PROMPT [17]	Point-to-point Top-down	Syntactic	Merging	No	Semi-automatic
FCA-Merge [14]	Point-to-point Bottom-up	Syntactic	Merging	Natural Language Document	Semi-automatic
ONION [18]	Point-to-point Top-down	Syntactic	Mapping & Merging	No	Semi-automatic
MOMIS [16]	Point-to-point Top-down	Syntactic	Mapping (similarity between nodes) & Merging	Thesaurus	Semi-automatic
CUPID [8]	Point-to-point Top-down	Syntactic	Mapping (similarity between nodes)	Thesaurus	Automatic (schema matching)
IF-based [13]	Mediated Bottom-up	Syntactic	Mapping	No	Automatic (Not yet shown)
GLUE [15]	Point-to-point Top-down	Syntactic & Semantic (domain constraints)	Mapping (similarity between nodes)	No	Semi-automatic
CTX-Match [9]	Point-to-point Top-down	Syntactic & Semantic (semantic relations)	Mapping (semantic relations between nodes)	WordNet	Automatic (identify relations)

The HCONE approach places human involvement at the early stages of the mapping/merging process. If this involvement leads to capturing the intended interpretation of conceptualisations, then the rest is a consistent, error-free merging process, whose results are subject to further human evaluation.

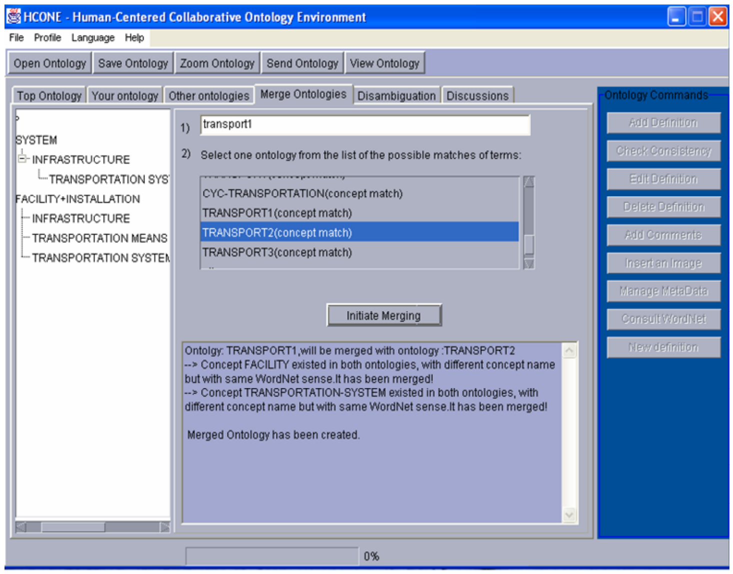


Fig. 4. HCONE-merge functionality. Merged concepts (e.g. FACILITY and INSTALLATION) are shown in the form Concept1+Concept2 (FACILITY+INSTALLATION) for presentation reasons

References

1. Uschold, M., Gruninger, M.: Creating Semantically Integrated Communities on the World Wide Web. Invited Talk, Semantic Web Workshop, WWW 2002 Conference, Honolulu, H1 (2002)
2. Kalfoglou, Y., Schorlemmer, M.: Ontology mapping: the state of the art. The Knowledge Engineering Review 18(1) (2003) 1–31
3. Kotis, K., Vouros, G.: Human-Centered Ontology Management with HCONE. In Proceedings of Ontologies and Distributed Systems Workshop, IJCAI-03 Conference, Acapulco, Mexico (2003)
4. Valarakos, A., Paliouras, G., Karkaletsis V., Vouros, G.: A name-Matching Algorithm for supporting Ontology Enrichment. In Proceedings of SETN’04, 3rd Hellenic Conference on Artificial Intelligence, Samos, Greece (2004)
5. Deerwester, S., Dumais, T. S., Furnas, W. G., Landauer, K. T., Harshman, R.: Indexing by Latent Semantic Analysis. Journal of the American Society of Information Science (1990)
6. Madhavan, J., Bern-stein, A. P., Domingos, P., Halevy, Y. A.: Representing and reasoning about mappings between domain models. In Proceedings of the 18th AAAI Canada (2002) 80–86
7. Giunchiglia F., Shvaiko, P.: Semantic Matching. In Proceedings of SemWeb-03 Conference, Workshop on Semantic Integration, Florida USA (2003)

8. Madhavan, J., Bern-stein, P. A., Rahm, E.: Generic schema matching with Cupid. VLDB Journal (2001) 49–58
9. Serafini, L., Bouquet, P., Magnini, B., Zanobini, S. : An Algorithm for Matching Contextualized Schemas via SAT. In Proceedings of CONTEX'03 (2003)
10. Uschold, M.: Where are the Semantics in the Semantic Web? AI Magazine, Vol. 24, No.3, Fall (2003)
11. Noy, F. N., Musen, A. M.: Evaluating Ontology-Mapping Tools: Requirements and Experience. In Proceedings of OntoWeb-SIG3 Workshop at the 13th International Conference on Knowledge Engineering and Knowledge Management, Siguenza, Spain, (2002) 1–14
12. Gangemi, A., Pisanelli, D. M., Steve, G.: An Overview of the ONIONS Project: Applying Ontologies to the Integration of Medical Terminologies. Data and Knowledge Engineering, Vol. 31 (1999) 183–220
13. Kalfoglou, Y., Schorlemmer, M.: Information-Flow-based Ontology Mapping. In Proceedings of the 1st International Conference on Ontologies, Databases and Application of Semantics, Irvine, CA, USA (2002)
14. Stumme, G., Mädche, A.: FCA-Merge: Bottom-Up Merging of Ontologies. In: B. Nebel (Ed.): Proceedings of 17th Intl. Joint Conference on Artificial Intelligence. Seattle, WA, USA (2001) 225-230
15. Doan, A., Madhavan, J., Domingos, P., Halvey, A.: Learning to map between ontologies on the semantic web. In Proceedings of 11^{th} International WWW Conference, Hawaii (2002)
16. Beneventano, D., Bergamaschi, S., Guerra, F., Vincini, M.: The MOMIS approach to Information Integration. IEEE and AAAI International Conference on Enterprise Information Systems, Setϋbal, Portugal, (2001)
17. Noy, F. N., Musen, A. M.: PROMPT: Algorithm and tool for automated ontology merging and alignment. In Proceedings of 7^{th} National Conference on AI, Austin, TX (2000)
18. Mitra, P., Wiederhold, G., Decker, S.: A Scalable Framework for Interoperation of Information Sources. In Proceedings of the 1st International Semantic Web Working Symposium, Stanford University, Stanford, CA (2001)