ONTOLOGY MATCHING BASED ON HYPERNYM, HYPONYM, HOLONYM, AND MERONYM SETS IN WORDNET

JungAe Kwak and Hwan-Seung Yong

Department of Computer Science and Engineering,
Ewha Womans University, Seoul, Korea
jakwak@ewhain.net, hsyong@ewha.ac.kr

ABSTRACT

Considerable research in the field of ontology matching has been performed where information sharing and reuse becomes necessary in ontology development. Measurement of lexical similarity in ontology matching is performed using synset, defined in WordNet. In this paper, we defined a Super Word Set, which is an aggregate set that includes hypernym, hyponym, holonym, and meronym sets in WordNet. The Super Word Set Similarity is calculated by the rate of words of concept name and synset's words inclusion in the Super Word Set. In order to measure of Super Word Set Similarity, we first extracted Matched Concepts(MC), Matched Properties(MP) and Property Unmatched Concepts(PUC) from the result of ontology matching. We compared these against two ontology matching tools – COMA++ and LOM. The Super Word Set Similarity shows an average improvement of 12% over COMA++ and 19% over LOM.

KEYWORDS

Ontology Matching, Property Unmatched Concept, Semantic Relationship Set, Super Word Set Similarity

1. INTRODUCTION

An ontology consists of concepts, attributes, and properties representing relationships between concepts. Ontology's properties can represent user-defined relationships as well as is-a and has-a relationships. Defining semantic relationships between concepts enables the development of a machine system that can automatically interpret and understand the meaning of concepts used in ontologies. For such automatic knowledge acquisition, ontologies have been widely studied and developed in a variety of domains, which increases the need for sharing and reusing available ontologies.

However, the task of sharing and reusing ontologies has many challenges. First, there is a problem related to polysemy and synonym. A word can have multiple meanings (i.e., polysemy) or there may be more than one word that means the same thing in a domain (i.e., synonyms), which causes difficulty in ontology matching. Second, polymorphism in expressions also brings about a difficulty in finding semantically related concepts. For example, an ontology might use noun forms of names and another ontology uses adjectival forms of names to indicate the same concept. Third, ontologies do not give a restriction on the count of a concept to represent a meaning. For example, in order to represent someone's name, an ontology simply has a concept–'Name', while another ontologies use two concepts 'FirstName' and 'LastName' [1]. Finally, both a concept and a property can be used to give a meaning. For example, an ontology can use the concept 'adviser' and another can use the property 'advisedBy' to indicate someone who gives advice [1]. Due to such difficulties, ontology matching has been the subject of

considerable research. In particular, as users increasingly demand a way to automatically exchange information and knowledge in applications, the automation of ontology matching becomes a research issue of great importance [2].

An Ontology matching is generally comprised of three phases of similarity measure – similarity measure between concepts, similarity measure between properties, and logical inference similarity measure; and there are four major dimensions of similarity, i.e., lexical similarity, structural similarity, instance similarity, and logical inference similarity. Existing ontology matching methods rely heavily on lexical similarity. Lexical similarity generally employs the character comparison method that focuses only on order of characters, irrespective of their meanings. However, this approach cannot solve the problem of polysemy and synonym described above. To address this drawback, lexical similarity in ontology matching is often measured using Synonym Set(synset) defined in WordNet. If names of concepts or properties of two ontologies are included in synset of WordNet, a pair of concepts or properties is determined, which results in ontology matching in measuring the lexical similarity.

In this paper, we propose an ontology matching method using the Super Word Set Similarity. The Super Word Set is an aggregate set that includes semantic relationships of words in WordNet (i.e. hypernyms, hyponyms, holonyms, and meronyms).

We performed a new form of concept matching based on the rate of including names of concepts and synset's words in the *Super Word Set*. The proposed method first extracts the *Matched Concepts*(MC) and the *Matched Properties*(MP) that are produced in the phase of similarity measure between concepts and in the phase of similarity measure between properties. Next, we extracted the *Property Unmatched Concepts*(PUC) that are not matched concepts among the domain concepts and range concepts of the MPs. The *Super Word Set Similarity* measure is then performed only between the PUCs, which eventually results in finding additional MCs. In this way, the proposed method can achieve a higher matching rate than the previous ontology matching methods.

This paper is organized as follows. Section 2 describes ontology matching tools and algorithms, similarity measures used in ontology matching, and existing research works using synset in WordNet. Section 3 presents the phase of ontology matching and the definitions of the terms used in this paper. Section 4 states the proposed ontology matching method using the *Super Word Set Similarity* measure, and the experiment results of the proposed method are given in Section 5. Finally, Section 6 presents conclusions and future areas of work.

2. RELATED WORK

2.1. Ontology matching tools and algorithms

LSD (Learning Source Description) [1] employs a multi-strategy learning approach. In the training phase, LSD chooses one learner among several base learners, such as Name Learner, Content Learner, and XML Learner, based on their confidence score. In the matching phase, ontology matching is performed using the learner chosen in the training phase, in order to improve matching accuracy.

GLUE [3], a system extended from LSD, performs semi-automatic ontology matching. GLUE creates a joint probability distribution table with neighbourhood-nodes, using the multi-strategy learning approach. In addition, GLUE supports ontology matching using Relaxation Labeling.

MOMIS (Mediator Environment for Multiple Information Source) [4] creates a lexical matrix using WordNet. A lexical matrix consists of rows storing word forms and columns storing word

meanings. Relations between words such as SYN (synonyms), BT (broader terms), NT (narrower terms), and RT (related terms) can be represented using the developed lexical matrix.

MAFRA(An Ontology Mapping Framework for the Semantic Web) [5] provides a distributed mapping process that consists of five horizontal and four vertical modules. It also performs ontology matching using concept and property bridges.

LOM (Lexicon-based Ontology Mapping) [6] proposes four phases of a lexical similarity measurement in ontology matching. We will describe these four phases in detail in Section 3.1.

COMA++ (Combining match algorithms) [7] is a well known ontology matching tool providing a graphical user interface. COMA++ adopts multi-match strategies to perform matching of relational schemas, W3C XSD, and OWL. The implemented multi-match strategies utilize fragment-based matching and reuse-oriented matching.

Unlike the previous ontology matching methods focusing on 1:1 mapping, RiMOM (Risk Minimization based Ontology Mapping) [8] automates the process of discoveries on 1:1, n:1, 1:null and null:1 mappings. Besides, RiMOM solves the problem of name conflicts in mapping process using thesaurus and statistical techniques.

ILIADS (Integrated Learning In Alignment of Data and Schema) [9] proposes an ontology matching algorithm based on logical inference. ILIADS integrates the previous ontology matching approach and the logical inference similarity measure to achieve better matching of ontologies.

2.2. Similarity measures in Ontology matching

In order to obtain ontology matching results, after each similarity in ontology matching is measured between concepts or properties, each similarity measurement is combined. The similarity measurements used in ontology matching are lexical similarity, structural similarity, instance similarity, and logical inference similarity [8, 9].

First, the lexical similarity measure has predominantly two approaches 1) character comparison through method of words used, name or description ontology's elements and 2) the use of WordNet. In the first approach, compound words and words with special characters or spaces are split into lexical tokens, and then the character comparison is performed. The other approach generally utilizes synset in WordNet. Existing works using synset in WordNet are described further in Section 2.3.

Second, ontology's concepts involve taxonomies and constraints. A taxonomy defines parent-child relationships of concepts and properties. In ontology matching, similarities between concepts can be calculated by analyzing ontologies' taxonomies and various constraints defined in association with concepts. This approach of comparing taxonomies of two ontologies and constraints in concepts is called the structural similarity measure.

Third, OWL(Web Ontology Language) defines any instance of a concept in RDF (Resource Description Framework). The instance similarity measures similarity in concepts by comparing instances of these concepts.

Fourth, OWL defines property characteristics such as FunctionalProperty, InverseOf, SymmetricProperty, TransitiveProperty, and InverseFunctionalProperty. Applications can use property characteristics to make logical inferences, and the reasoned logical inferences are then used to discover matched concepts. This approach is called the logical inference similarity measure [9].

2.3. Ontology Matching Using WordNet's synset

WordNet is a thesaurus with defined semantic vocabulary set. Bundles of the same semantic words are called synset, meaning synonym set and are identified by their synsetID. Synsets of WordNet are connected with semantic relationships. For example, nouns are linked in terms of hypernym, hyponym, coordinate term, holonym, and meronym relationship; verbs have the relationships of hypernym, troponym, entailment, and coordinate terms [10]. Existing ontology matching methods using WordNet mostly use synsets in WordNet. It matches by the *equivalentClass* relationship between two concepts if two concept's names are included of the same synset of WordNet.

Lin proposed a method that measures the similarity between two words and computes their cosine coefficient using WordNet. The similarity here is computed by the rate of the number of words in synsets to which the word of interest belongs, to the total number of words in WordNet [11].

Gouli proposed COMPASS (COncept MaP ASSessment tool) which uses WordNet to give reflective feedback to learners concerning matching results. If a learner used a hyponym or meronym of the expected word, the matching would be still considered correct. Otherwise, the tool recommends using a more general concept [12].

In this paper, following an analysis of previous studies about ontology matching, we define the *Property Unmatched Concept*. In the next two sections, we present how the ontology matching method can discover additional matched concepts by measuring the *Super Word Set Similarities* between the *Property Unmatched Concepts* are also presented in detail.

3. ONTOLOGY MATCHING

3.1. Ontology Matching Process

Figure 1. Ontology matching process

As shown in Figure 1, two ontologies, represented as O_1 and O_2, are needed for ontology matching. Based on input ontologies O_1 and O_2, an ontology matching tool first performs concept matching by measuring the similarity between concepts. Next, property matching is carried out by measuring the similarity between properties. The logical similarity measure is then performed based on the results of previous concept matching and property matching. Lastly, the final matching results are produced after passing through the refinement process.

LOM [6] proposes a lexical similarity measure in ontology matching. LOM performs a lexical similarity by 4 phases. In the first phase, whole term matching, LOM finds an exact name string match. In second phase, word constituent matching, stop words such as 'a', 'the', 'of', 'in', etc. are dropped from multi-word terms. Remaining words for each term are compared through exact string matching. After performing this phase, 'meeting-place' and 'place-of-meeting' can be matched. In third phase, Synset matching, if two words are included in synset in WordNet, two words are matched. Through this phase, 'auto-care' and 'car-maintenance' can be matched. In fourth phase, Type matching, LOM performs ontology matching using SUMO(the Suggested

Upper Merged Ontology) [13,14] and MILO(the Mid-level Ontology) [15]. Existing ontology matching tools using a lexical similarity perform first three phases.
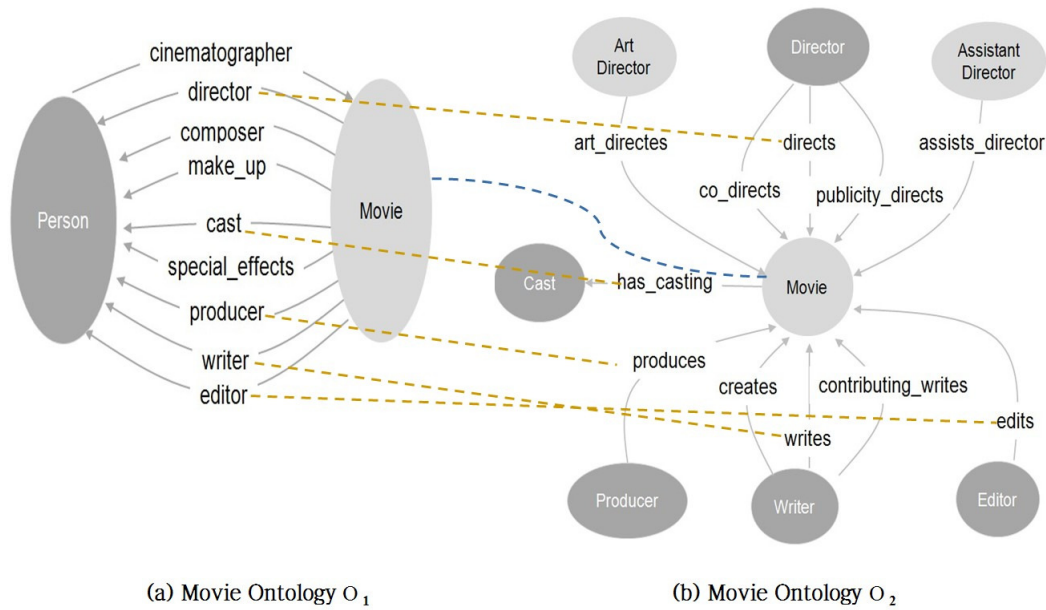

Figure 2. Movie ontologies O_1, O_2 and matching results. Dotted lines mean a matching

OWL distinguishes between two main categories of properties, datatype properties and object properties. Datatype properties declare data type defined in XML Schema. Object properties define semantic relations between concepts. In two concepts semantically linked via an object property, one becomes a domain concept when the other is a range concept [16]. Object properties are mainly used in this paper. An object property is represented below.

Ontology_name.(Domain concept_name, *Property_name*, Range concept_name)

For example, movie ontology O_1 in Figure 2 has a property represented as **O_1.(Movie,*director*,Person)** where *director* is an object property, **Movie** is a domain concept, and **Person** is a range concept.

Figure 2 shows the final matching results of two movie ontologies O_1 and O_2, produced through the matching processes in Figure 1. The matched concepts **O_1.Movie** and **O_2.Movie** are found in the phase of similarity measure between concepts. In the phase of similarity measure between properties, five pairs of properties are matched including **O_1.(Movie,*director*,Person)** and **O_2.(Director,*directs*,Movie)**. Table 1 lists the matching results of the two movie ontologies depicted in Figure 2.

Table 1. Matching results of movie ontologies O_1 and O_2

matching	Ontology O_1	Ontology O_2
Concept matching	Movie	Movie
Property matching	director	directs
	cast	has_casting
	producer	produces
	writer	writes
	editor	edits

Analyzing the concepts appearing in the properties O_1.(**Movie**,*director*,**Person**) and O_2.(**Director**,*directs*,**Movie**) shows that the domain concept O_1.**Movie** in the property O_1.(**Movie**,*director*,**Person**) and the range concept O_2.**Movie** in the property O_2.(**Director**,*directs*,**Movie**) have been already matched in concept matching phase.

The range concept O_1.**Person** in the property O_1.(**Movie**,*director*,**Person**) and the domain concept O_2.**Director** in the property O_2.(**Director**,*directs*,**Movie**) are not yet matched. Let's analyze O_1.**Person** and O_2.**Director**. The synset in WordNet of 'person' word contains {person, individual, someone, somebody, mortal, soul} and the synset in WordNet of 'director' word contains {director, manager, managing director}. O_1.**Person** and O_2.**Director** are not matched because these two words do not contain the same synset of WordNet. In this paper, we propose an ontology matching approach in order to search new concept matching using hypernyms, hyponyms, holonyms, and meronyms in WordNet along with synset in WordNet between O_1.**Person** and O_2.**Director**.
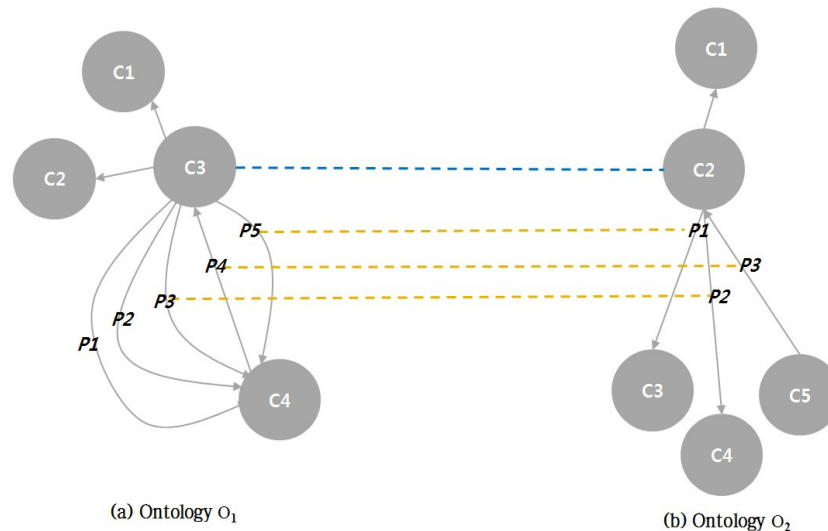

Figure 3. An ontology matching diagram. O_1 and O_2 ontology examples and matching results. Dotted lines mean a matching

3.2. Terminology

In this paper, a concept matched in the phase of similarity measure between concepts is called a *Matched Concept*(MC), represented as **O**.$\underline{C}$. Whereas, **O.C** represents an unmatched concept. Similarly, a property matched in the phase of similarity measure between properties is called a *Matched Property*(MP), which is denoted as **O**.(**$\underline{DC,p,RC}$**). An unmatched property is represented as **O**.(**DC***,p,***RC**). Note that MPs apply only to object properties of the MCs. In Figure 3, **O$_1$.C3** and **O$_2$.C2** are matched by means of the relationship *equivalentClass*, and are denoted as **O$_1$.$\underline{C3}$** and **O$_2$.$\underline{C2}$**. There are three pairs of the MPs corresponded by means of the relationship *equivalentProperty*, i.e., **O$_1$.(C3***,p5,***C4)** and **O$_2$.(C2***,p1,***C3)**, **O$_1$.(C4***,p4,***C3)** and **O$_2$.(C5***,p3,***C2)**, and **O$_1$.(C3***,p3,***C4)** and **O$_2$.(C2***,p2,***C4)**. According to the notation described above, the first pair is denoted as **O$_1$.($\underline{C3,p5,C4}$)** and **O$_2$.($\underline{C2,p1,C3}$)**.

Among the domain concepts and range concepts of the MPs, those that are not matched are defined as a *Property Unmatched Concept*(PUC), and denoted as **O**.$\underline{C}$. In Figure 3, the PUCs are **O$_1$.C4**, **O$_2$.C3**, **O$_2$.C4**, and **O$_2$.C5**. They are denoted as **O$_1$.$\underline{C4}$**, **O$_2$.$\underline{C3}$**, **O$_2$.$\underline{C4}$**, and **O$_2$.$\underline{C5}$**.

Synonym set	Description
person, individual, someone, somebody, mortal, soul	a human being; "there was too much for one person to do
person	a human body (usually including the clothing); "a weapon was hidden on his person
person	a grammatical category of pronouns and verb forms; "stop talking about yourself in the third person

(a) Synonym set in WordNet of person noun

Semantic relationship of a noun	Description
hypernyms	person is a kind of….
hyponyms	… is a kind of person
coordinate Terms	
holonyms	person is a part of….
moronyms	part of person

(b) List of semantic relationship of a noun

Figure 4. Example of WordNet's synonym sets and semantic relationships

WordNet groups words into sets of synonyms, and provides a short, general description of each set of synonyms. Figure 4.a shows that the word 'person' belongs to three synonyms sets. The first synonym set contains the words {person, individual, someone, somebody, mortal, soul}. In addition, words found in synset of WordNet are interlinked by various semantic relationships. For example, WordNet's nouns are interlinked in terms of hypernym, hyponym, coordinate term, holonym, and meronym, as listed in Figure 4.b.

This paper defines the *Super Word Set* that contains all the words and their descriptions in association with WordNet's semantic relationships. Note that WordNet's synsets are not included in the *Super Word Set* because they are already used for matching in the phase of similarity measure between concepts.

$$SWS(\text{Super Word Set}) = \sum w \in \{\text{hyperset}\} + \sum w \in \{\text{hyposet}\} +$$
$$\sum w \in \{\text{holoset}\} + \sum w \in \{\text{meroset}\}$$

In the equation above, hyperset is a set of hypernyms; hyposet is a set of hyponyms; holoset is a set of holonyms; and meroset is a set of meronyms. Each of these sets is composed of a set of words and their descriptions.

4. SUPER WORD SET SIMILARITY

4.1. Extraction of the Property Unmatched Concept

The MCs and MPs should be determined first in order to search the PUCs. In the matching results of the movie ontologies presented in Figure 2 and Table 1, O_1.**Movie** and O_2.**Movie** are the MCs. Among all the object properties of the MCs, there are five pairs of MPs as shown below:

O_1.(**Movie**,*director*,**Person**) and O_2.(**Director**,**directs**,**Movie**),

O_1.(**Movie**,*cast*,**Person**) and O_2.(**Movie**,*has_casting*,**Cast**),

O_1.(**Movie**,*producer*,**Person**) and O_2.(**Producer**,*produces*,**Movie**),

O_1.(**Movie**,*writer*,**Person**) and O_2.(**Writer**,*writes*,**Movie**),

O_1.(**Movie**,*editor*,**Person**) and O_2.(**Editor**,*edits*,**Movie**).

Then, among the domain concepts and the range concepts of these five pairs of MPs, those that are not previously matched become candidates for a PUC. To determine whether or not these candidates should become a real PUC, the property matching rate of each MC is calculated by dividing the number of MPs by the number of all existing object properties of the MC. If the calculated property matching rate is greater than a given critical value α (α ∈ [0, 100]), the candidates are chosen as a PUC.

In Figure 2, the MC O_1.**Movie** has 5 MPs and its total number of object properties is 9. Thus, O_1.**Movie** has the property matching rate 55% (5/9*100= 55%). The MC O_2.**Movie** has 5 MPs and 11 object properties, therefore its property matching rate is 45% (5/11*100 = 45%). When the critical value is 0.45 (α = 0.45), the candidates of O_1.**Movie** and O_2.**Movie** are chosen as a PUC, i.e., O_1.**Person**, O_2.**Writer**, O_2.**Editor**, O_2.**Director**, O_2.**Producer**, and O_2.**Cast.**

4.2. Super Word Set Similarity Measure

The *Super Word Set Similarity* is measured to discover new matched concepts based on PUCs. The *Super Word Set Similarity* measure proposed in this paper makes use of the *Super Word Set* presented in Section 3.2 and WordNet's synsets. The *Super Word Set Similarity* measure of two ontologies O_i and O_j is defined by a percentage of included synset words of the PUC of O_i in the PUC's *Super Word Set* of O_i. The *Super Word Set Similarity* measure can be computed in two ways, i.e., from O_i ontology to O_j ontology, and vice versa.

$$sim_{sws}(O_i.C_1, O_j.C_2) = \frac{n(synset(O_j.C_2) \in SWS(O_i.C_1))}{n(SWS(O_i.C_1))}.$$

where $O_i.C_1$ and $O_j.C_2$ are a pair of PUCs in two ontologies.

Table 2. A result table of the Super Word Set Similarity of the Property Unmatched Concept

O_1.PRC	O_2.PRC	$O_1 \rightarrow O_2$			$O_2 \rightarrow O_1$		
		n(SWS)	n(synset)	sim_{sws}	n(SWS)	n(synset)	sim_{sws}
Person	Writer	410	3	0.4%	48	8	33 %
	Editor	410	4	0.4%	18	8	17 %
	Director	410	12	0%	26	8	23 %
	Producer	410	4	0%	18	8	44 %
	Casting	410	7	0.4%	120	8	3 %

Table 2 shows the results of measuring the *Super Word Set Similarity* between the identified PUCs of the movie ontologies in Figure 2 - i.e., O_1.**Person**, O_2.**Writer**, O_2.**Editor**, O_2.**Director**, O_2.**Producer**, and O_2.**Cast**.

The range of the calculated *Super Word Set Similarity* values in Table 2 is rather broad, from 0% to 44%. Thus, a critical value for the *Super Word Set Similarity* is used. If the calculated *Super Word Set Similarity* of a pair of PUCs is greater than a given critical value β ($\beta \in [0, 100]$), that pair of PUCs is matched, and the PUCs become new MCs. Based on experiments, the critical value β is set to 20% ($\beta = 20\%$). In Table 2, among 10 pairs of PUCs, 3 pairs have the *Super Word Set Similarity* measure greater than the given critical value; they are sim_{sws}(O_2.**Writer**, O_1.**Person**), sim_{sws}(O_2.**Director**, O_1.**Person**), and sim_{sws}(O_2.**Producer**, O_1.**Person**). These PUCs are thus determined as new MCs.

4.3. Determination of Semantic Relationship

The semantic relationship of a pair of MCs newly discovered by the *Super Word Set Similarity* measure also needs to be determined. A critical value β is defined to determine the types of semantic mapping between new MCs. Based on experiments, γ is set to 80% ($\gamma=80\%$). The logic to determine the semantic relationship type based on γ is shown below.

$$
\begin{aligned}
&\text{if} \ \ \text{sim}_{sws}\,(O_i.C_1, O_j.C_2) < \ \gamma \ \text{then} \\
&\qquad (O_i.C_1, \textit{subClass}, O_j.C_2) \\
&\text{else} \\
&\qquad (O_i.C_1, \textit{equivalentClass}, O_j.C_2)
\end{aligned}
$$

where $O_i.C_1$ and $O_j.C_2$ are a pair of new MCs in two ontologies.

In Table 2, sim_{sws}(O_2.**Writer**, O_1.**Person**) has the *Super Word Set Similarity* measure 33% (less than the critical value 80%), therefore the pair's semantic relationship type would be *subClass*, denoted as O_2.(**Writer**,*subClass*,**Person**).

Figure 5 shows the ontology matching process in which the phase of *Super Word Set Similarity* measure is added. The added phase is placed in between the phase of similarity measure between properties and the phase of logical inference similarity measure.
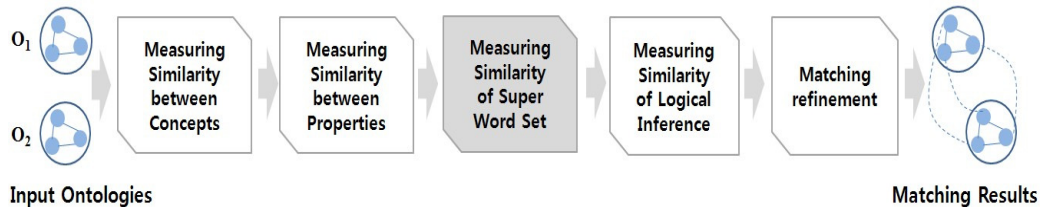

Figure 5. Ontology matching process added the phase of Super Word Set Similarity measure

5. EXPERIMENTAL EVALUATION

This section explains the metrics used to assess the accuracy and efficiency of ontology matching, and compares the effectiveness of the proposed method with that of COMA++ and LOM. Ontologies in the domain of movie and food are used in the experiments.

5.1. Measures for Experimental Evaluation

In this paper, the Precision and Recall are used to evaluate the proposed method.
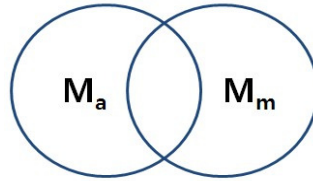

Figure 6. Precision and Recall

This section explains the metrics used to assess the accuracy and efficiency of ontology matching, and compares the effectiveness of the proposed method with that of COMA++ and

In Figure 6, M_a indicates a set of MCs and MPs via an ontology mapping tool, which is called automated matching. However, M_m is a set of MCs and MPs that are matched manually by a domain expert, and is therefore called manual matching. The formula to compute Precision and Recall are shown below.

Precision is the ratio of the intersection of an automatically matched set and a manually matched set to the automatically matched set.

$$\text{Precision} = M_a \cap M_m / M_a$$

Recall is the ratio of the intersection of an automatically matched set and a manually matched set to the manually matched set.

$$\text{Recall} = M_a \cap M_m / M_m$$

5.2. Experimental results

In the experiments of the proposed method, the PUCs were identified based on the MCs and MPs matched in the phase of concept matching and the phase of property matching. The *Super Word Set Similarity* was then measured on the extracted PUCs, and new MCs were determined according to the measured *Super Word Set Similarity*.

Figure 7 shows the final matching results of two food ontologies O_1 and O_2, produced through the matching processes in Figure 1. As shown in Figure 7, **O_1.Meal** and **O_2.MEAL**, **O_1.MealCourse** and **O_2.MEAL-COURSE** are extracted as the MCs and **O_1.(Meal,*course*,MealCourse)** and **O_2.(MEAL,*COURSE*,MEAL-COURSE)**, **O_1.(MealCourse, *hasDrink*,PotableLiquid)** and **O_2.(MEAL-COURSE,*DRINK*,JUICE)** or , **O_2.(MEAL-COURSE,*DRINK*,WINE)** are extracted as the MPs. Based on those matching results, we can extract **O_1.PotableLiquid, O_2.JUICE** and **O_2.WINE** as PUCs.
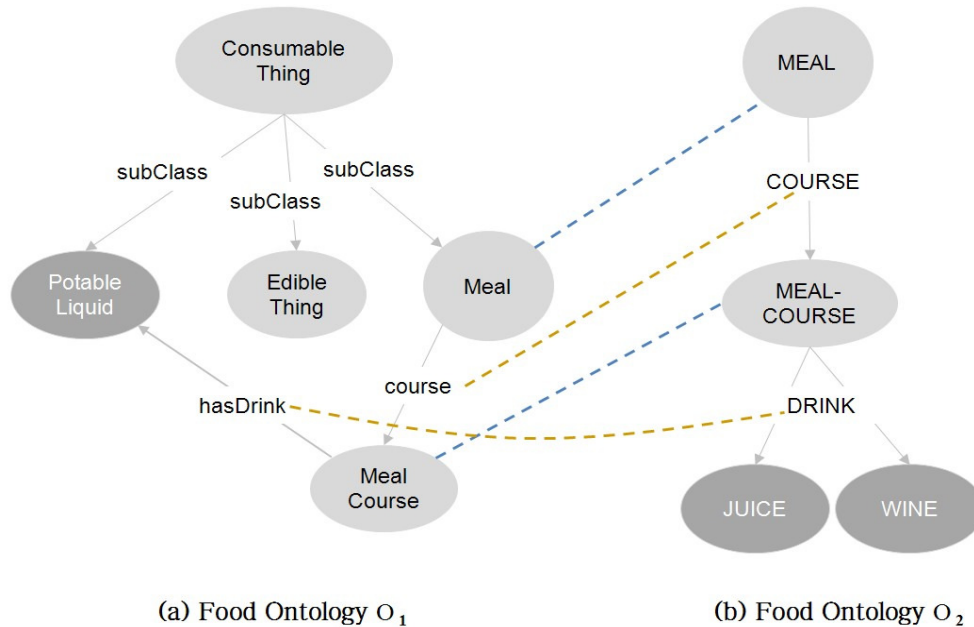
(a) Food Ontology O$_1$ (b) Food Ontology O$_2$

Figure 7. Food ontologies O$_1$, O$_2$ and matching results. Dotted lines mean a matching.

Table 3. The number of concepts, properties and results of manual matching

Ontology	Ontology O$_1$		Ontology O$_2$		n(manual matching)
	n(Concept)	n(Property)	n(Concept)	n(Property)	
Movie Ontology	3	25	332	99	116
Food Ontology	63	4	166	10	70

Table 3 shows the number of concepts, properties and manual matching of the movie and food ontologies used in the experiments. Movie ontology O$_1$ has properties eight times higher than the number of concepts, which indicates that it is property oriented ontology. On the other hand, movie ontology O$_2$ has the number of concepts more than three times higher than the number of properties, so it is presumed to be a concept oriented ontology. Food ontologies O$_1$ and O$_2$ are both concept oriented ontology. Considering that food ontology O$_2$ has the number of concepts more than two times higher than food ontology O$_1$, one can Figure out that O$_2$ is built in a more detailed way than O$_1$.
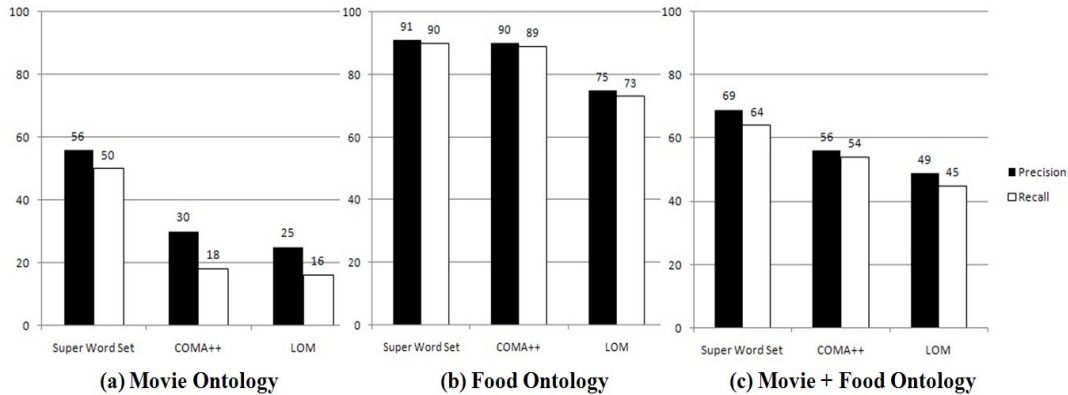
Figure 8. Experimental result

In the studies, the proposed method based on the *Super Word Set Similarity* is compared with COMA++ and LOM in order to examine the accuracy and efficiency of their ontology matching. Figure 8 shows the Precision and Recall rates of the two compared methods with regard to movie ontologies, food ontologies, and a combination of movie and food ontologies.

In the movie ontologies, both Precision and Recall rates are under 60% as the property oriented ontology O_1 and the concept oriented ontology O_2 are matched. On the other hand, in food ontologies O_1 and O_2 are concept oriented ontology, high Precision and Recall rates are attained and are over 70%.

The proposed *Super Word Set Similarity* method has the advantage of finding new MCs based on the PUCs that are derived from the previously matched concepts and properties. As shown in Figure 8, the proposed *Super Word Set Similarity* method performs better than COMA++ in movie ontology matching. This is because many PUCs were extracted from the MCs and MPs that are found in the phases of concept matching and property matching. This implies that PUCs are extracted most effectively when a concept oriented ontology and property oriented ontology are matched.

On the other hand, the proposed method and COMA++ yield similar results when two concept oriented ontologies are matched, as shown in the matching results of the food ontologies in Figure 8. This is because the number of PUCs extracted from the previously MCs and MPs was relatively small.

6. CONCLUSIONS AND FUTURE WORKS

This paper has presented the *Super Word Set Similarity* measure that makes use of WordNet to discover new MCs in ontology matching. The proposed ontology matching method based on the *Super Word Set Similarity* takes the matching results of the phase of similarity measure between concepts and the phase of similarity measure between properties in order to find new MCs. The experiments carried out demonstrate that the proposed method result in an average improvement of 12% over COMA++ and 19% over LOM.

Unlike previous ontology matching tools that utilize only WordNet's synonym sets, the proposed method employs semantic relationships set of WordNet - i.e., *Super Word Set* including collectively WordNet's hypernyms, hyponyms, holonyms, and meronyms is used for ontology matching. In the proposed method, additional MCs are newly discovered based on the inclusion rate of synset words over Super Word Set. Currently, the proposed *Super Word Set Similarity* method is applied for concepts that have a one-word noun. In future work, how to address compound nouns, adjectives and verbs will be studied. That is, this research will be

extended to build a *Super Word Set* that includes all types of words so that the proposed method is an applicable to a wider range of word types.

ACKNOWLEDGEMENTS

REFERENCES

[1]   AnHai Doan, Pedro Domingos, and Alon Halevy, (2003) "Learning to Match the Schemas of Data Sources: A Multistrategy Approach," *Machine Learning*, Vol. 50, No.3, pp 279-301.

[2]   Namyoun Choi, Il-Yeol Song, and Hyoi Han, (2006) "A Survey on Ontology Mapping," *ACM SIGMOD Record*, Vol. 35, No, 3, pp 34-41.

[3]   AnHai Doan, Jayant Madhavan, Robin Dhamankar, Pedro Domingos, and Alon Y. Halevy, (2003) "Learning to match ontologies on the Semantic Web," *The VLDB Journal*, Vol. 12, No 4, pp 303-319.

[4]   Domenico Beneventano, Sonia Bergamaschi, Francesco Guerra, and Maurizio, (2003) "Synthesizing an Integrated Ontology," *IEEE Internet Computing*, pp 42-51.

[5]   Alexander Maedche, Boris Motik, Nuno Silva, and Raphael Volz, (2002) "MAFRA - A MApping FRAmework for Distributed Ontologies in the semantic Web," *Proc. of the Workshop on Knowledge Transformation for the Semantic Web (KTSW 2002)*, pp 60-68.

[6]   John Li, (2004) "LOM: A Lexicon-based Ontology Mapping Tool," *Proc. of the Performance Metrics for Intelligent Systems Workshop (PerMIS. '04)*, pp 1-5.

[7]   David Aumueller, Hong-Hai Do, Sabine Massmann, and Erhard Rahm, (2005) "Schema and ontology matching with COMA++ ," *Proc. of the ACM SIGMOD International Conference on Management of Data*, pp 906-908.

[8]   Jie Tang, Juan-Zi Li, Bangyong Liang, Xiaotong Huang, Yi Li and Kehong Wang, (2006) "Using Bayesian decision for ontology mapping," *Journal of Web Semantics*, Vol. 4, No. 4, pp 243-262.

[9]   Octavian Udrea, Lise Getoor, and Renée J. Miller, (2007) "Leveraging data and structure in ontology integration," *Proc. of the ACM SIGMOD International Conference on Management of Data*, pp 449-460.

[10]   WordNet, http://wordnet.princeton.edu/

[11]   Patrick Pantel and Dekang Lin, (2002) "Discovering Word Senses from Text," *Proc. of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp 613-619.

[12]   Harry Kornilakis, Maria Grigoriadou, Kyparisia A.Papanikolaou, and Evangelia Gouli, (2004) "Using WordNet to Support Interactive Concept Map Construction," *Proc. of the IEEE International Conference on Advanced Learning Technologies (ICALT'04)*, pp 600-604.

[13]   Ian Niles and Adam Pease, (2001) "Toward a Standard Upper Ontology," *Proc. of the International Conference on Formal Ontology in Information Systems (FOIS-2001)*, pp 2-9.

[14]   Adam Pease, and Ian Niles, (2003) "Linking Lexicons and Ontologies: Mapping WordNet to the Suggested Upper Merged Ontology," *Proc. of the IEEE International Conference on Information and Knowledge Engineering*, pp 412-416.

[15]   Ian Niles, and Allan Terry, (2004) "The MILO: A general purpose, mid-level ontology," *Proc. of the International Conference on Information and Knowledge Engineering (IKE'04)*, pp 15-19.

[16]   Grigoris Antoniou and Frank van Harmelen, (2004) *A Semantic Web Primer*, The MIT Press, Massachusetts, pp 118.

Authors

JungAe Kwak is a Ph.D. candidate at Computer Science and Engineering at Ewha Womans University, Republic of Korea. She received the M.S. degree in Information Science and Engineering at Ewha Womans University. Her current major research interests include semantic web, ontology matching, and ontology modelling.

Hwan-Seung Yong   is a Professor of Computer Science and Engineering at Ewha Womans University, Republic of Korea since 1995. He received the B.S., M.S. and Ph.D. degrees in Computer Engineering from Seoul National University. He has five years of industrial experience as Researcher at ETRI (Electronics and Telecommunications Research Institute) and served as a visiting scientist at the IBM T.J. Watson Research Center. He is member of KIISE (The Korean Institute of Information Scientists and Engineers) and ACM. His current major research interests include agile methods, data mining, multimedia database and ontology search.

Short Biography