

Dealing with Uncertain Entities in Ontology Alignment using Rough Sets

Sadaqat Jan, Maozhen Li, Hamed Al-Raweshidy, Alireza Mousavi and Man Qi

Abstract — Ontology alignment facilitates exchange of knowledge among heterogeneous data sources. Many approaches to ontology alignment use multiple similarity measures for mapping entities between ontologies. However, it remains a key challenge in dealing with uncertain entities for which the employed ontology alignment measures produce conflicting results on similarity of the mapped entities. This paper presents OARS, a Rough sets based approach to ontology alignment which achieves a high degree of accuracy in situations where uncertainty arises because of the conflicting results generated by different similarity measures. OARS employs a combinational approach and considers both lexical and structural similarity measures. OARS is extensively evaluated with the benchmark ontologies of the Ontology Alignment Evaluation Initiative (OAEI) 2010, and performs best in the aspect of recall in comparison with a number of alignment systems while generating a comparable performance in precision.

Index Terms— Ontology alignment, Rough sets, semantic matching, semantic interoperability, knowledge engineering.



1. INTRODUCTION

ONTOLOGIES facilitate exchange of knowledge between heterogeneous data sources. An ontology is an explicit formal specification of the terms in a domain and relations among them [1][2]. As the number of ontologies grows, it is now common to have different ontologies for a single domain [3]. Ontology designers might have distinct objectives in mind while developing an ontology depending on their application demands. It is essential to utilize well defined parts from all the available ontologies for a particular domain to achieve the best results of knowledge sharing. Ontologies can be heterogeneous in various forms including *terminological heterogeneity* and *conceptual heterogeneity*. These heterogeneities must be dealt with in an ontology alignment process which plays a vital role in semantic interoperability between applications. The alignment process aligns the semantically related entities defined in heterogeneous ontologies which have been developed for a similar domain.

In recent years, a number of alignment systems have been proposed which includes automatic, semi-automatic, application specific and general purpose systems as analyzed and

reviewed in several aspects [3][4][5]. The schema matching techniques [6] are also intensely examined by the research community as the ontology alignment process primarily requires the identification of the correspondences between semantically related entities. During an automatic alignment process, entities are selected for mapping when a certain level of semantic correspondence is found leaving the dissimilar entities unmapped. Most of the existing ontology alignment approaches compare the similarities using more than one elementary technique and then the results of these techniques are aggregated using a variety of aggregation strategies [7][8][9]. The combination of structural and lexical techniques produces an overall better similarity of a concept defined in an ontology. Each individual matching technique is treated as a matcher and the results of all the matchers can be aggregated in different ways to finalize the alignment process. These aggregation methods may employ weighted average techniques or probabilistic methodologies to calculate the probability of an entity in a source ontology being similar to an entity in a target ontology. However, the real issue arises when a combinational method turns out to be uncertain with the entities which are neither completely similar nor dissimilar because of the conflicting results generated by individual matchers. Thus, finding uncertain entities and dealing with such uncertain entities are an even more complicated task than finding only similar or dissimilar entities in the ontology alignment process. Such uncertain entities are becoming more prevalent when partial information about a concept is available in one ontology as compared to the information available on the same concept in another ontology.

This paper presents OARS, a novel ontology alignment approach to dealing with uncertain entities in ontology mapping. OARS builds on Rough sets [10] to compute the similarities of ontology entities in an alignment process. In OARS, the entities are first matched through three elemen-

- Sadaqat Jan is with the Computer Software Engineering Department at Khyber Pakhtunkhwa University of Engineering and Technology, Mardan Campus, Pakistan. E-mail: Sadaqat_Jan@hotmail.com
- Maozhen Li is with the School of Engineering and Design, Brunel University, Uxbridge, UB8 3PH, UK. E-mail: Maozhen.Li@brunel.ac.uk. He is also associated with the Key Laboratory of Embedded Systems and Service Computing, Ministry of Education, Tongji University, China.
- Hamed Al-Raweshidy is with the School of Engineering and Design, Brunel University, Uxbridge, UB8 3PH, UK. E-mail: Hamed.Al-Raweshidy@brunel.ac.uk.
- Alireza Mousavi is with the School of Engineering and Design, Brunel University, Uxbridge, UB8 3PH, UK. Email: Ali.Mousavi@brunel.ac.uk.
- Man Qi is with the Department of Computing, Canterbury Christ Church University, Canterbury, Kent, CT1 1QU, UK. Email: Man.Qi@canterbury.ac.uk.

Manuscript received (insert date of submission if desired). Please note that all acknowledgments should be placed at the end of the paper, before the bibliography.

tary matchers which are based on structures, strings and linguistics respectively. The entities for which the individual matchers cannot reach a consistent mapping decision on their similarity will be considered as uncertain entities to be processed by the Rough sets classification in OARS. The unmapped entities generated by the three matchers are defined as the attributes of the corresponding elements of Rough sets. OARS classifies a set of elements based on the available attributes and computes the accuracy ratio of Rough sets classification to reach a mapping decision on uncertain entities.

OARS has been extensively evaluated using the benchmark ontologies of the Ontology Alignment Evaluation Initiative (OAEI) 2010¹, and it performs best in the aspect of recall when comparing with a number of OAEI participating alignment systems. In addition, OARS produces a comparable performance in the aspect of precision.

It is worth noting that OARS is extended from our initially proposed alignment system [11] and is further evaluated with the three groups of the benchmark data sets. More importantly, the significance of using Rough sets as an aggregation method is also evaluated in this paper. Furthermore, we have integrated OARS into our previously developed SemFARM [12][13], a framework which provides an efficient search mechanism for file annotation and retrieval on mobile devices connected through Bluetooth. The integration of OARS enables SemFARM to utilize the knowledge of multiple ontologies when searching for a file on resource-limited devices in a network environment which leads to high accuracy in file retrieval.

The rest of the paper is structured as follows. Section 2 reviews related work on ontology alignment. In Section 3, we describe in detail the similarity measures and the matching process of OARS. Section 4 presents the Rough sets classification which deals with uncertain entities in ontology mapping. Section 5 evaluates the performance of OARS using the benchmark ontologies of OAEI 2010. Section 6 integrates OARS into the SemFARM framework for enhanced file retrieval on mobile devices, and Section 7 concludes the paper.

2. RELATED WORK

In recent years, a significant research has been conducted to deal with ontology alignment. In this section, we review the related work on traditional approaches which have not considered the issue of uncertainty in the mapping process and the emerging approaches that have considered this issue.

2.1 Traditional Approaches to Ontology Alignment

These research efforts mainly follow two approaches. One approach uses single matchers to match ontology entities by comparing their label (name) information with corresponding synonyms. Normally WordNet² is exploited in such an approach. For example, the similarity function employed by Rodriguez and Egenhofer [14] is based on a matching process which uses synonym sets along with other

available information from ontology specifications [15]. Other features of such a lexicon are also exploited to find more relationships between the entities such as hypernym, hyponym, meronym and holonym [16][17]. Single matcher based alignment systems only work well in aligning those ontologies which have similar internal and external structures. Using structural matching techniques, the comparison is made between the entities based on their structural positions in ontologies, set of properties, domain, data-types and cardinality. GMO [18] is an example of structural matchers which takes a set of matched pairs as external output in the matching process and uses bipartite graphs to compare the structural similarity of different ontologies. The V-Doc matcher [19] measures the context of domain entities in terms of their meanings in the Vector Space Model. However, any alignment technique in isolation like GMO or V-Doc is not adequate enough for an accurate mapping result. For this reason, OARS incorporates string, linguistic and structural-based matchers.

Another approach to ontology alignment aggregates a number of single matchers. For example, RiMOM [8] uses multiple matchers to discover lexical and structural similarities between entities and exploits Bayesian decision theory in order to map them. The basic matchers which are considered as separate strategies compare the taxonomy, constraints, descriptions, names, instances and name-paths in the mapping process. The user input is also allowed to improve the mapping in the alignment process. The enhanced version of RiMOM [20] exploits most of the available ontological knowledge by using these strategies via a strategy selection technique and combines all the similarity values using a sigmoid function, and then initiates an alignment refinement algorithm to finalize the alignment process. However, the parameter settings in RiMOM are highly depended on the preprocessing step where two similarity factors are compared in ontologies and the weights are then assigned to different factors for combining the final results. This means that if two ontologies have more structural similarities, a higher value will be assigned to the weight of structural similarity in combining the final result. Therefore, the mapping of those ontology entities which have other similarities will suffer because the same parameters will be used for all the entities. In OARS, we use Rough sets classification for each individual entity and the mapping decision is made on the entity bases which do not affect the overall decision of other mappings.

Falcon-AO [9] uses a combination of linguistic, structural and partition based matchers in the mapping process. Falcon-AO is based on the alignment work of V-Doc, I-Sub [21] and GMO. Falcon-AO requires a similarity combination strategy in order to combine the similarity value generated by each matcher. A set of coordination rules are used to reduce structural heterogeneity as a pre-mapping process. The alignment results are returned to determine the equality and subsumption relationships between classes and properties. Isaac et al. [61][62] evaluated the effectiveness of Falcon in thesaurus merging which is mainly attributed to its lexical component. However, using linguistic similarity, Falcon-AO does not differentiate between data-type properties and object properties while in OARS we use the linguistics

¹ <http://oaei.ontologymatching.org/2010/>

² <http://wordnet.princeton.edu/>

tic matcher for classes and properties separately. This avoids any chances of mapping a class-entity in one ontology with a property-entity in another ontology.

ASMOV [22] is an automatic ontology matching tool which uses both structural and lexical matchers to calculate similarity for ontology integration. ASMOV automates the alignment process using the weighted average of measurements of similarity and obtains a pre-alignment iteratively which is then verified for semantic inconsistencies. The semantic verification process examines the correct correspondences and incompleteness using predefined inferences. It requires more than one execution to finalize the mapping result and the results of the intermediate iterative executions are employed to refine the subsequent processing phases of alignment. However, the verification process does not provide efficient rules for unverified alignments.

The SOBOM algorithm [23] finds the anchors in the first step and uses Semantic Inductive Similarity Flooding (SISF) to flood similarity among concepts. Then it utilizes the results of the SISF mechanism to find the relationships between alignments. The SOBOM algorithm heavily depends on the precision of anchors returned by the linguistic matcher, i.e. the overall alignment performance will be degraded if the matcher misses an anchored concept.

AgrMaker [24] uses a three layer architecture in which a number of concepts and structural based matchers are included. It combines the results using a local confidence quality measure. AgrMaker mainly focuses on providing rules to combine different mapping sets rather than identifying the matching itself. CODI [25] uses Markov logic based probabilistic alignment which transforms the alignment process into a Maximum-a-Posteriori optimization problem. It combines lexical similarity measures with schema information for matching entities in the alignment process. The performance of CODI is highly dependent on the pre-alignment mappings.

TaxoMap [26] takes into account the labels and sub-class descriptions in ontologies for alignment and employs the Partition based Block Matching algorithm [27] which allows the use of predefined equivalence mappings to partition the ontologies into pairs of possible mappings. MapPSO [28] considers ontology alignment as an optimization problem and employs the Discrete Particle Swarm Optimization algorithm [29] for solving the problem. Using the MapPSO approach all particles are updated and adjusted iteratively for the best representative particles in the swarm. However, the performance of MapPSO depends on the selection of high quality matchers and aggregators.

Although the aforementioned systems have certain merits in ontology alignment, unlike OARS, they have not considered the uncertainty issues during the alignment process as emphasized in Section 1.

2.2 Emerging Approaches to Ontology Alignment

It is worth noting that only a few ontology alignment systems have considered uncertainty during the mapping process. For example, the alignment system proposed in [60] deals with uncertain entities in such a way that it employs Dempster-Shafer theory to aggregate the mapping results generated by individual matchers. Dempster-Shafer theory is

also employed in the work presented in [33] to deal with uncertainty in ontology mapping. Sváb and Svátek [31] employed Bayesian networks to model mapping methods and aggregations of the mapping results. To produce accurate mapping results, the conditional probability tables in the constructed Bayesian networks need to be well trained through a learning process. Pan et al. [32] presented a Bayesian network based approach to dealing with uncertainty in ontology mapping. The source and the target ontologies are first translated into Bayesian networks. Then the mappings of the concepts (entities) between the two ontologies are processed as evidential reasoning between the two translated Bayesian networks. This approach is based on an assumption that each concept is associated with sufficient and high quality exemplars during a learning process. It is worth noting that Garruzzo and Rosaci [63] presented a method to cluster semantically homogeneous agents. A set of explanations are employed for agents to resolve uncertain terms in communication. However, the efficiency of this method is highly depended on the completeness of the explanation set. Moreover, this method demands a sufficient number of semantic negotiation steps among the agents in communication.

OARS builds on Rough sets to deal with uncertainty in ontology alignment. Different from the aforementioned approaches which are based on Dempster Shafer theory and Bayesian networks, Rough sets theory does not need any preliminary or additional information about data which means that Rough sets theory is objective in information processing as highlighted by Li et al. [34].

3. SIMILARITY MEASURES

The overall ontological heterogeneities have been categorized in many aspects and presented in detailed reviews [30][35][36][37]. There are mainly two types of heterogeneity namely *semantic* and *terminological* heterogeneity. *Semantic* heterogeneity occurs due to various reasons like using different axioms or disparity in modeling the same concept. *Terminological* heterogeneity emerges when using synonyms or different names for the same entity in different ontologies. In order to deal with most types of ontological heterogeneities, OARS follows a combinational approach and uses lexical and structural matchers along with WordNet as an external resource to compute the semantic similarity between entities. The three individual matchers that are employed in OARS are based on existing techniques or with some minor modifications which are explained in the subsequent sections. To align two ontologies, a *source ontology* O and a *target ontology* O' , OARS uses three matchers to compute the similarity between the entities of O and O' :

- The string based matcher is used to find the similarity between the named classes and entities.
- The WordNet-based linguistic matcher is used to compare semantic similarity.
- In the structural based matcher, the super-classes and sub-classes are compared taking into account the constraints to find the similarity of object and data properties of the classes.

3.1 String-based Similarity

In string-based similarity calculation, the entities are considered as strings regardless of their structures or other associated properties. A string normalization process is performed after the basic comparison of entity names. Both entity strings are converted to lower-case and punctuations, dashes and blank characters are eliminated. The normalization process is crucial in string comparisons. For example, “*MasterThesis*”, “*Master-Thesis*” and “*Master thesis*” are normalized to “*masterthesis*”. A number of techniques are proposed to calculate the string similarities using the characteristics of measurements. These techniques include substring distance [37][38], Levenstein [39], Jaro-Winkler [40][41], Needleman-Wunsch [42] and n-gram similarity [43][44]. A good survey on string distance calculation can be found in [45].

Stoilos et al. [21] proposed the *Smoa* string metric which is based on the intuitions about similarities presented in [46]. *Smoa* computes string similarity based on strings commonalities as well as their differences. The *Smoa* metric is calculated by subtracting the sum of differences and *winkler* similarity from the commonalities of strings. The commonalities are calculated using a substring metric.

Let Sim_strng denote the string similarity between entities e_i and e'_i , then $Sim_strng(e_i, e'_i)$ can be calculated using equation (1).

$$Sim_strng(e_i, e'_i) = Smoa(e_i, e'_i) \quad (1)$$

To calculate a substring metric between two strings, a process to find and remove the largest common substring is continued until no further common substrings can be found. The lengths of these substrings are then added and scaled with the lengths of strings. The differences used in *Smoa* are computed with the lengths of unmatched strings. The *Smoa* measurement is used in OARS as a string-based matcher.

3.2 Linguistic Similarity

Linguistic based similarities are computed using external resources like language dictionaries, thesauri or specific databases. Such similarities are useful when string-based similarities are difficult to find between ontology entities especially when synonyms are used for the same concept in ontologies. For example, the names “*brochure*” and “*booklet*” refers to the same concept but the string-based similarity between them is low enough (*which is 6, using the Levenshtein distance*) to be ruled out for selection as a mapping candidate. The WordNet [47] is a lexical database which provides a repository of lexical items defined as a set of semantic vocabulary. In WordNet, different meanings of the same concept are grouped together as sets of synonyms (*synsets*) in terms of nouns, verbs, adjectives and adverbs. Synsets are interlinked in a hierarchical structure using various conceptual-semantic and lexical relationships. For example, nouns have relationships such as hypernym, hyponym, holonym, meronym among the words. Similarly verbs are linked through relationships of hypernym, troponym, entailment, and coordinate terms. Now consider the same example of the two entity names “*brochure*” and “*booklet*”. They will be selected as good candidates for mapping in WordNet where the *brochure*, *folder*, *leaflet* and *pamphlet*

are defined as synonyms.

For linguistic similarity, context-based measures can also be employed. For example, Sahami et al. [64] defined a new kernel function to measure the semantic similarity between pairs of short text snippets by utilizing context vectors. Banerjee et al. [65] measured the semantic relatedness of concepts by utilizing the hierarchies of concepts presented in lexical databases like WordNet. Similarly, Patwardhan and Pedersen [66] utilized the co-occurrence information along with the WordNet definitions to build gloss vectors corresponding to each concept and assigned numeric scores to pairs of concepts by computing the cosine of the angle between their respective gloss vectors.

OARS employs WordNet to exploit the information encoded in the names and labels of the ontology entities. Using WordNet, we consider the synonyms, hyponyms, hypernyms and antonyms of entities.

Let

- $Sim_lin(w_i, w'_i)$ be the linguistic similarity between words w_i and w'_i ,
- Σ be the external resource (*WordNet*),
- $s(w_i)$ be the set of synonyms,
- $h(w_i)$ be the set of hyponyms and hypernyms,
- $t(w_i)$ be the set of antonyms of w'_i ,

The linguistic similarity of two words w_i and w'_i can be computed using equation (2).

$$Sim_lin(w_i, w'_i) = \begin{cases} 1 & \text{if } w'_i \in s(w_i) \\ 0.5 & \text{if } w'_i \in h(w_i) \\ 0 & \text{if } w'_i \in t(w_i) \end{cases} \quad (2)$$

The similarity relationships of hyponyms and hypernyms are set to 0.5 and are further computed in structure matching using equations (3), (4), (5) and (6). For words which are synonyms and antonyms they will be considered as *similar* and *dissimilar* respectively. One possible drawback of using resources like WordNet is that a number of possible matches might be found for the same concept [48]. To solve this problem, OARS employs three types of structural information of the possible matches of the entities which will be described in the following section.

3.3 Structural Similarity

The structural information plays a vital role in situations where both the linguistic and string based similarity matches between two ontology entities are proved to be insufficient or incomplete. For example, Sánchez et al. [67] utilized the ontology structures to improve the accuracy of the modelled taxonomical knowledge. In [68], Sánchez et al. also reviewed a number of structural similarity measures including the similarity measures based on super-classes. Sub-classes are also exploited in structure based similarity measures between ontologies [69][70]. Similarly, OARS exploits the information on super-classes and sub-classes of the ontologies to compute the structural similarities of the entities. The main intuitions behind the structural similarity in OARS

are given below:

- If two classes from different ontologies have similar super-classes in hierarchy, it is likely that they define the same concept.
- If two classes from different ontologies have similar sub-classes in hierarchy, it is likely that they define the same concept.
- If two classes from different ontologies have similar properties, it is likely that they define the same concept.
- It is likely that two entities having any combination of two or all of the three above mentioned similarities share the similar concept.

The structure similarity of two entities e_i and e'_i from ontologies O and O' respectively can be computed taking into account the similarities between the super-classes, sub-classes and properties of the two entities.

Let

- $Sim_hsp(e_i, e'_i)$ be the structural similarity between the super-classes of entities e_i and e'_i ,
- $K_{sup}(e_i)$ be the set of super classes of entity e_i ,
- $K_{sup}(e'_i)$ be the sets of super classes of entity e'_i ,
- $|K_{sup}(e_i)|$ be the cardinality of $K_{sup}(e_i)$,
- $|K_{sup}(e'_i)|$ be the cardinality of $K_{sup}(e'_i)$,

we have

$$Sim_hsp(e_i, e'_i) = \frac{1}{2} \left(\frac{|(K_{sup}(e_i) \cap K_{sup}(e'_i))|}{|K_{sup}(e_i)|} + \frac{|(K_{sup}(e_i) \cap K_{sup}(e'_i))|}{|K_{sup}(e'_i)|} \right) \quad (3)$$

Let

- $Sim_hsb(e_i, e'_i)$ be the structural similarity between the sub-classes of entities e_i and e'_i ,
- $K_{sub}(e_i)$ be the set of sub classes for entity e_i ,
- $K_{sub}(e'_i)$ be the sets of sub classes for entity e'_i ,
- $|K_{sub}(e_i)|$ be the cardinality of $K_{sub}(e_i)$,
- $|K_{sub}(e'_i)|$ be the cardinality of $K_{sub}(e'_i)$,

we have

$$Sim_hsb(e_i, e'_i) = \frac{1}{2} \left(\frac{|(K_{sub}(e_i) \cap K_{sub}(e'_i))|}{|K_{sub}(e_i)|} + \frac{|(K_{sub}(e_i) \cap K_{sub}(e'_i))|}{|K_{sub}(e'_i)|} \right) \quad (4)$$

The similarity between the properties of entities also plays an important role in determining the overall similarity of two entities in different ontologies.

Let

- $Sim_pr(e_i, e'_i)$ represent the similarity between the properties of entities e_i and e'_i ,

- $Pr(e_i)$ be the set of properties of entity e_i ,
- $Pr(e'_i)$ be the set of properties of entity e'_i ,
- $|Pr(e_i)|$ be the cardinality of $Pr(e_i)$,
- $|Pr(e'_i)|$ be the cardinality of $Pr(e'_i)$,

then, we have

$$Sim_pr(e_i, e'_i) = \frac{1}{2} \left(\frac{|(Pr(e_i) \cap Pr(e'_i))|}{|Pr(e_i)|} + \frac{|(Pr(e_i) \cap Pr(e'_i))|}{|Pr(e'_i)|} \right) \quad (5)$$

Finally, the overall structural similarity $Sim_strc(e_i, e'_i)$ of two entities is computed by the average of the three structural matchers using equation (6).

$$Sim_strc(e_i, e'_i) = 1/3 (Sim_hsp(e_i, e'_i) + Sim_hsb(e_i, e'_i) + Sim_pr(e_i, e'_i)) \quad (6)$$

4. USING ROUGH SETS FOR SIMILARITY AGGREGATION

Rough sets theory is based on the indiscernibility relation of objects with respect to the available information which partitions the universe into sets of similar objects called elementary sets [49]. The elementary sets can further be used to build knowledge on the real or abstracted world where the use of indiscernibility relation leads to information granulation. Rough sets theory has proved to be a useful mathematical technique for analysing object descriptions. It assumes that every object of the universe is associated with a certain amount of information, represented by some attributes which express the descriptions of objects [50][51]. The detailed discussion about applications of Rough sets in knowledge discovery and data mining is given in [52].

The concept of objects and their attributes in Rough sets is exploited in OARS to deal with uncertainties during the mapping process of ontology alignment when the results of the individual matchers do not give a definitive indication on whether the entities are similar or dissimilar. Using Rough sets, the similarity results of the individual matchers are considered as the attributes of elements for classification which is further used to determine the similarities between the elements based on their attribute values.

Let

- U be a set of unmapped entities in a target ontology, $U = \{e_1, e_2, e_3, \dots, e_n\}$,
- F be a set of matching factors which represents the coverage or importance of individual matching results, $F = \{f_1, f_2, f_3\}$,
- X be a subset of U .

Let $[x]_F$ denote a set of entities which have similarities among them with regard to given matching factors. The lower and upper approximations of the set X are defined as follows.

Let

- $\underline{F}(X)$ represent the lower approximation of the set X with respect to F , the set of matching factors. Then $\underline{F}(X)$ is a set of entities that certainly belong to X , as defined by expression (7).

$$\underline{F}(X) = \{x \mid [x]_F \subseteq X\} \quad (7)$$

- $\bar{F}(X)$ represent the upper approximation of the set X with respect to F . Then $\bar{F}(X)$ is a set of entities that may possibly belong to X , as defined by expression (8).

$$\bar{F}(X) = \{x \mid [x]_F \cap X \neq \emptyset\} \quad (8)$$

The accuracy ratio of Rough sets classification for lower and upper approximations of the set X can be computed using equation (9).

$$\alpha_F(X) = \frac{|\underline{F}(X)|}{|\bar{F}(X)|} \quad (9)$$

The ratio of the accuracy will be in the range of $0 \leq \alpha_F(X) \leq 1$. For a selected entity from the source ontology, OARS computes the similarity to each unmapped entity in the target set U . For entities for which the three individual matchers, i.e. $Sim_strng(e_i, e'_i)$, $Sim_lin(w_i, w'_i)$ and $Sim_strc(e_i, e'_i)$ do not find exact matches between them, the similarity results generated by these matchers will be classified by Rough sets for each of unmapped elements. For entities in the set X which are definable [10], [53] with respect to F , these entities will be considered for mapping when the accuracy ratio of Rough sets classification is 1 based on equation (9). The set F defines three matching factors (f_1, f_2, f_3) as follows for assigning a confidence degree in Rough sets classification.

- f_1 represents the value of $Sim_strng(e_i, e'_i)$, as defined in equation (1).
- f_2 represents the mean value of $Sim_hsp(e_i, e'_i)$ and $Sim_hsb(e_i, e'_i)$ as defined in equations (3) and (4) respectively.
- f_3 represents the value $Sim_pr(e_i, e'_i)$ as defined in equation (5).

The linguistic matcher ($Sim_lin(w_i, w'_i)$) is not considered in computing the three matching factors because it only produces a fixed value of 0.5 based on equation (2) for uncertain entities in mapping which is not suitable for Rough sets classification. The similarity of two entities is computed from four aspects, i.e. string similarity, super-class similarity, sub-class similarity, and property similarity. Each aspect of similarity is computed with a weight of 25% which means that f_1 or f_3 only evaluates 25% of the total similarity represented by the set F , while f_2 evaluates 50% of the total similarity represented by the set F . To maximize the set of entities to be classified by Rough sets, the values of the matching factors are normalized to the nearest decimal values before computing the accuracy ratio of Rough sets classification.

Algorithm 1 shows the pseudo code of the Rough sets classification in mapping entities. Line 1 is used to assign the three matching factors. Lines 2-6 are used to select the entities for alignment based on the accuracy ratio of Rough Sets classification. Lines 7-10 are used to assign a confidence degree to the mapped entities.

Algorithm 1: Mapping entities using Rough sets classification.

```

Input:  $E = \{e_1, e_2, e_3, \dots, e_m\}$ , a set of unmapped entities
          from the source ontology;
           $E' = \{e'_1, e'_2, e'_3, \dots, e'_n\}$ , a set of unmapped entities
          from the target ontology;
           $F_1 = F$ ,  $F = \{f_1, f_2, f_3\}$ , a set of matching factors;
           $F_2 \subset F$ ,  $F_2 = \{f_1, f_2\}$ ;
           $F_3 \subset F$ ,  $F_3 = \{f_2, f_3\}$ ;
Output: align  $(e_i, e_j, c)$  where  $c$  is a confidence degree;
1: For  $k = 1$  to 3;
2:   For  $i = 1$  to  $m$ ;
3:     For  $j = 1$  to  $n$ ;
4:       compute  $\alpha_F$  as defined in equation (9);
5:       If  $\alpha_F = 1$ , then
6:         align  $(e_i, e'_j)$ ;
7:       If  $F_k = F_i$  then
8:          $c = 1$ ;
9:       Else
10:         $c = 0.75$ ;
11:       Endif
12:     Endif
13:   Endfor
14: Endfor

```

To further illustrate the use of Rough sets in OARS to determine the similarities between ontology entities, we present an example as shown in Fig.1. We assume that both cases have 5 unmapped entities namely e'_1, e'_2, e'_3, e'_4 and e'_5 in the target ontology. The three matching factors are given against each target entity after comparing with the entity e_i in the source ontology. In this example, we only compare e_i with e'_1 and e'_4 respectively.

We present two separate cases namely *Case-1* and *Case-2*. *Case-1* is presented to demonstrate the similarity calculation between the source and target entities which is explicitly based on two factors f_1 and f_2 . While *Case-2* is presented to demonstrate the similarity calculation taking into account the three factors f_1, f_2 and f_3 .

Consider *Case-1* as given in Fig. 1:

- for $X = \{e_i, e'_1\}$, with respect to $F = \{f_1, f_2\}$, the $\bar{F}(X) = \{\{e_i, e'_4\}, \{e'_1, e'_3\}\}$, and $\underline{F}(X) = \emptyset$ indicating that e_i and e'_1 are indefinable based on the given results and left unmapped.
- for $X = \{e_i, e'_4\}$, with respect to $F = \{f_1, f_2\}$, both $\bar{F}(X)$ and $\underline{F}(X) = \{e_i, e'_4\}$, and the $\alpha_F(X) = 1$ indicating that e_i and e'_4 are considered for mapping. The confidence degree value of 0.75 is assigned to the

mapping relationship because the set F contains two matching factors in this case.

Consider *Case-2* as given in Fig.1:

- for $X = \{e_i, e'_1\}$, with respect to $F = \{f_1, f_2, f_3\}$, the $\bar{F}(X) = \{e_i, e'_4\}$, and $\underline{F}(X) = \emptyset$ indicating that e_i and e'_1 are indefinable based on the given results and left unmapped.
- for $X = \{e_i, e'_4\}$, with respect to $F = \{f_1, f_2, f_3\}$, both $\bar{F}(X)$ and $\underline{F}(X) = \{e'_1, e'_4\}$, $\alpha_F(X) = 1$ indicating e_i and e'_4 are definable with respect to F , and the two entities are considered for mapping with a confidence degree of 1.

As discussed earlier, the Rough sets classification classifies the objects based on specific attributes. Similarly, in this example, using case-1, the objects (in this case, objects are entities e_i and e'_4) are considered for mapping based on the attributes (in this case attributes are f_1 and f_2).

The alignment process in OARS is shown in Fig.2 which starts with pre-processing to normalize the names of the ontology entities as discussed in Section 3.1. OARS then uses the three individual matchers to compute the similarity values of the entities between the source and the target ontologies using equations (1), (2) and (6). If an exact similarity is found by any individual matcher, the entities are selected for mapping and a confidence degree of 1 is assigned. For uncertain entities, they will be fed into Rough sets classification for further computation. After the mapping process, OARS verifies that any entity in the source ontology is not mapped to more than one entity in the target ontology and vice versa. If such a mapping is found, OARS selects the mapping with the higher confidence degree before producing the final alignments.

wide range of features regarding the strengths and weaknesses of the existing matchers. The reference alignments are also available for tests which have been aligned manually by OAEI and considered as correct alignments.

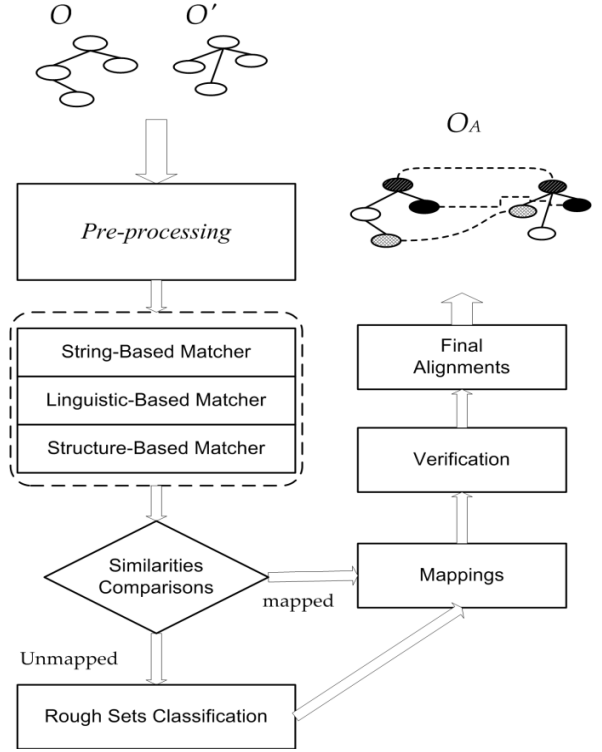


Fig.2. OARS alignment process.

5.1 Benchmark Data Sets

The OAEI 2010 benchmark data sets include a number of ontologies with varied levels of complexities. These ontologies are built from one OWL ontology on the bibliography topic. The base ontology is test-101 which is considered as a reference ontology, containing 33 named classes, 24 object properties, 40 data properties and 76 individuals of which 20 of them are named while the rest are anonymous.

The descriptions of these tests are given in Table 1, mainly containing three groups - *simple tests* (1xx), *systematic tests* (2xx) and *real-life ontologies* (3xx). The 1xx group has 4 ontologies with minor variations. Ontologies in the systematic tests (group 2xx) have been built to test the ability of the alignment systems when specific information is eliminated from the ontologies. The eliminated information may include the following:

- Classes are replaced with several classes, expanded or flattened.
- The entity names are replaced with synonyms, strings from other languages than English or even some random strings.
- Comments at different levels are translated into other foreign languages than English or suppressed at all.
- Properties are suppressed or their restrictions on classes are discarded.
- Instances are suppressed.

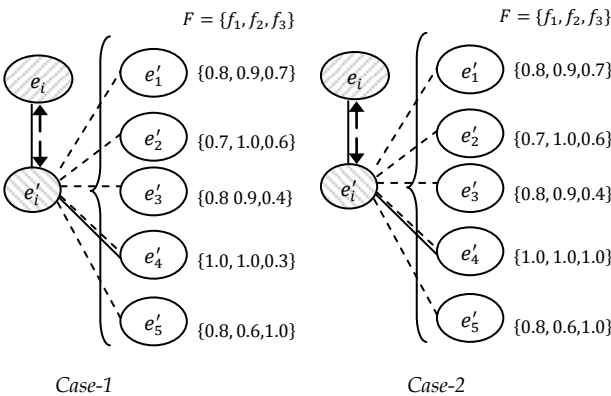


Fig.1. An example of Rough sets classification.

5. EVALUATION

We have implemented OARS using the Java programming language and the alignment API [54] to input source and target ontologies and to generate alignment results. We used the benchmark ontologies of the OAEI alignment campaign 2010 to evaluate the performance of OARS. These benchmark tests offer various sets of ontologies to evaluate a

- Specialized hierarchies are expanded, suppressed or flattened.

Furthermore, ontologies in group 3xx are real world ontologies which are provided by different institutions and left unchanged in the benchmark data sets.

Table 1. The descriptions of the benchmark data sets³.

Test sets	Descriptions
101-104	The hierarchical structure is similar Entity name is same or totally different
201-210	The hierarchical structure is similar Different linguistic used in some levels
221-247	Different in structure Label linguistic is similar
248-266	Hierarchical structure and linguistics are different
301-304	Real world ontologies

5.2 Evaluation Measures

We use *precision*, *recall* and *F-measure* to evaluate the accuracy of OARS in ontology alignment. Precision and recall are the most widely accepted and well-known measures in the research areas of information retrieval [55], [56] and ontology alignment [57].

Let A_d be the set of produced alignments, A_t be the complete set of accurate alignments. The precision, recall and *F-measures* can be defined using equations (10), (11) and (12) respectively.

$$Prec = \frac{|A_d \cap A_t|}{|A_d|} \quad (10)$$

$$Rec = \frac{|A_d \cap A_t|}{|A_t|} \quad (11)$$

$$F - measure = \frac{2 \times Prec \times Rec}{Prec + Rec} \quad (12)$$

5.3 Experimental Results

This section presents the performance evaluation of OARS in a number of scenarios. The evaluation of similarity aggregation is presented to underline its effect on the results of overall performance in ontology alignment. A comparison of OARS with other existing alignment systems is also outlined in this section. Critical analyses are presented to highlight the advantages and limitations of OARS. The alignment process in OARS is totally automatic and hence no user intervention is involved in any tests during the alignment process.

5.3.1 Similarity Aggregation

To evaluate the performance of OARS comprehensively, we have formulated several test scenarios using the benchmark data sets and the evaluation criteria defined by (10), (11) and (12). The main purposes of these test scenarios are to assess:

- The efficacy of the individual similarity matchers,
- The effectiveness of various combinations of these individual matchers, and

- The effect of Rough sets classification in aggregating the results of the individual matchers.

We designed four scenarios of which each scenario uses different combinations of matchers to aggregate the final mapping results. For this purpose, we implemented four algorithms separately in the alignment system, namely $A1$, $A2$, $A3$ and $A4$ as defined by expressions (13), (14), (15) and (16) respectively. The details of these four algorithms are given below.

- $A1$ represents the method where ontology alignment is derived using the mean value of the results returned by the string and linguistic matchers,

$$A1 = (Sim_strng(e_i, e'_i) + Sim_lin(w_i, w'_i))/2 \quad (13)$$

- $A2$ represents the method where alignment is derived using the mean value of the results returned by the structural and linguistic matchers,

$$A2 = (Sim_strc(e_i, e'_i) + Sim_lin(w_i, w'_i))/2 \quad (14)$$

- Similarly, $A3$ uses the mean value of the results generated by the string and structural based matchers for alignment,

$$A3 = (Sim_strng(e_i, e'_i) + Sim_strc(e_i, e'_i))/2 \quad (15)$$

- Finally, $A4$ uses the mean value of the results generated by the string, linguistic and structural matchers for alignment,

$$A4 = (Sim_strng(e_i, e'_i) + Sim_lin(w_i, w'_i) + Sim_strc(e_i, e'_i))/3 \quad (16)$$

We selected the group 3xx of the test ontologies from the benchmark data sets because it contains the real world ontologies as described in Section 5.1. Fig.3 shows the comparison results of the methods $A1$, $A2$, $A3$ and $A4$.

The set of ontologies in group 3xx have more string similarities than structural and linguistic similarities in comparison with the reference ontology. From Fig.3 it is also evident that the algorithms ($A1$, $A3$ and $A4$) using the string-based matcher show better results in *F-measure* than $A2$ which does not use the string based matcher. This also indicates the significance of a single matcher in aligning the ontologies with favorable features. The linguistic matcher does not perform well in group 3xx ontologies because it cannot deal with some entities with prefix text such as “abstract”=“hasAbstract”, “volume”= “hasVolume” and “copyright”=“hasCopyright” using the WordNet synsets. Such results degrade the overall mapping performance of other matchers when a mean value of all the matchers is taken in aggregation. In Fig.3, the $A3$ algorithm does not consider the result of the linguistic matcher producing a better *F-measure* value than other algorithms.

We also compared the performance of OARS with that of the $A4$ method using the ontologies of the group 3xx. As shown in Fig.4, there is a significant improvement in the

³ <http://oaei.ontologymatching.org/2010/benchmarks/index.html>

performance of OARS as compared to *A4* in the three aspects. The precision, recall and *F-measure* values of *A4* are 0.805, 0.582 and 0.675 respectively while for OARS these values are 0.862, 0.845 and 0.83 respectively. The overall improvement achieved by OARS in *F-measure* is 22.96% over the *A4* method.

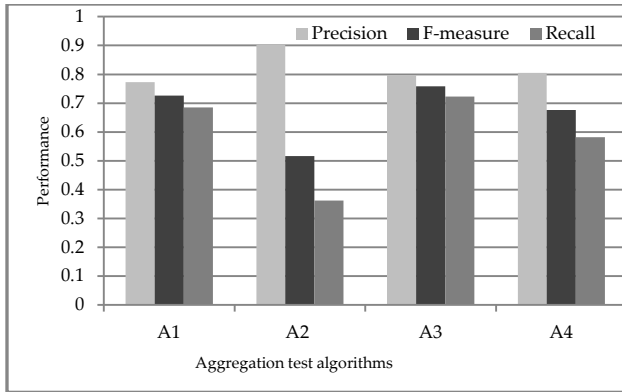


Fig.3. The performance of the four aggregation algorithms.

These evaluation results further fortify that no single matcher is sufficient enough to achieve high accuracy in ontology alignment. More importantly, simply aggregating the results of individual matchers by taking a mean value is not only insufficient but can also degrade the overall mapping performance when some matchers present low similarity values.

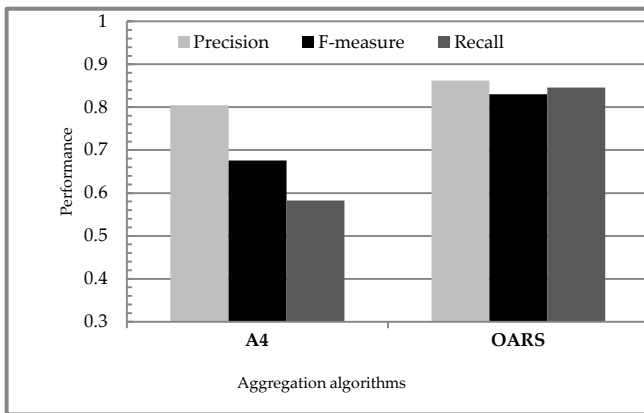


Fig.4. A comparison of aggregation algorithms.

5.3.2 Normalization in Rough Sets Classification

To select the most appropriate value for normalizing the results of the individual matchers for Rough sets classification, we performed various tests considering the normalization values of 50, 33.33, 25, 20, 10 and 5. These tests were performed on group 2xx of the benchmark data set. Fig.5 shows that OARS achieves the highest recall value using the normalization value of 50, but on the other hand it gives the lowest precision value. Similarly, using the normalization value of 5, OARS produces the highest precision value but gives the lowest recall value. We used the value of 10 in OARS for normalization as it produces the best *F-measure* value.

5.3.3 Comparing OARS with Existing Alignment Systems

This section evaluates OARS in comparison with a number of alignment systems which participated in the OAEI 2010 campaign using the benchmark data sets of group 1xx, group 2xx and group 3xx respectively. The published results can be found in [58].

Group 1xx

Almost all of the alignment systems in comparison achieved perfect results for ontologies in group 1xx in terms of precision and recall. However, one exception is TaxoMap which achieved a low recall value of 0.34. The good performance of these alignment systems in these tests is mainly attributed to the fact that the ontologies in group 1xx have highly similar entities. As there is no structural heterogeneity among these ontologies, only the string and linguistic based matchers were used in OARS for ontology alignment in group 1xx.

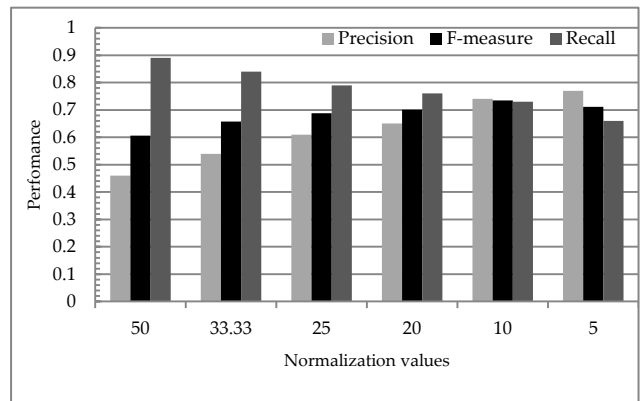


Fig.5. Normalization evaluation.

Group 2xx

Most of the ontologies in group 2xx were aligned correctly by OARS using the linguistic matcher which relies on the WordNet for dealing with synonyms (for example in test 205). The string based matcher also performed well on string heterogeneities. The linguistic matcher proved to be effective in ontologies where linguistics were used, for example in test ontologies 201, 202 and 248-266. Furthermore, ontologies with only structural changes were also tackled successfully in OARS because when this information was suppressed, the linguistic or string similarities were still available in the ontologies. We found that the most challenging alignment task was to deal with those ontologies in which both structural and labels modifications were made. In the tests on group 2xx, OARS achieved the best recall among the alignment systems as shown in Fig.6 because of its capability in dealing with uncertain entities in ontology mapping. It is worth noting that other alignment systems such as ASMOV, AgrMaker and RiMOM also achieved high recall values of 0.89, 0.83 and 0.84 respectively.

Group 3xx

There are 4 real world ontologies in group 3xx which have the blend of obscurities found in group 2xx data. In the tests on group 3xx data, as there is little structural information available in these ontologies, e.g. ontology 302,

OARS mainly relied on string and linguistic based matchers in aligning the ontologies in group 3xx. The test results of this group are plotted in Fig.7 which shows that ASMOV produces the best result in recall followed by OARS with a value of 0.86.

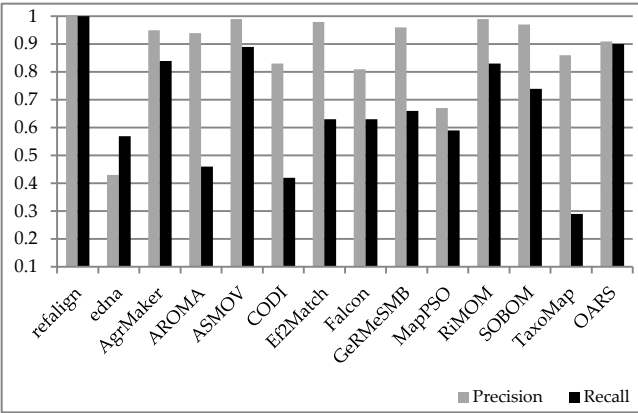


Fig.6. Evaluation results on group 2xx.

It is worth noting that the performance of OARS in precision is comparable to that of the other alignment systems which is reflected in both Fig.6 and Fig.7 respectively.

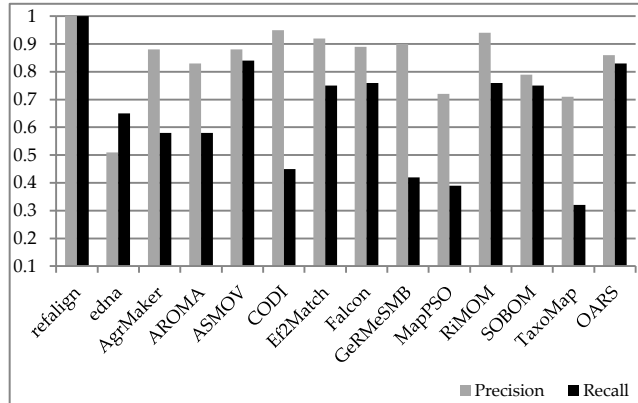


Fig.7. Evaluation results on group 3xx.

6. INTEGRATING OARS INTO SEMFARM

As described in Section 1, we have implemented SemFARM for file annotation and retrieval on mobile devices. To exploit the ontology alignment capabilities of OARS in SemFARM, a new search module was implemented in SemFARM. The annotation process automatically annotates the files with three basic attributes and two user entered fields. The meta-data is automatically parsed and stored in XML structured document as described in [12], [13] and [59]. Fig.8 shows the overall process of the SemFARM search module in which the input file queries are answered after merging two existing ontologies. When multiple OWL ontologies are found on the query answering system, they are first aligned and these alignments are then converted to associated axioms in order to utilize the alignments as a single ontology. For this purpose, initially one of the render classes OWLaxiomsRendererVisitor in the ontology align-

ment API package [54] was used which provides OWL axioms for expressing the relationships of equivalence, subsumption and exclusion. This renders the generated alignments as a merged ontology of the two input ontologies. Once the merged ontology is acquired, the ontology model and the RDF model are bound together to form an inference model. The RDF model is automatically created from the XML document by the XML to RDF converter module as shown in Fig. 8. Finally, the file-search query is answered by navigating the inference model for query-words. The list of the names of files is then returned to the corresponding sending device.

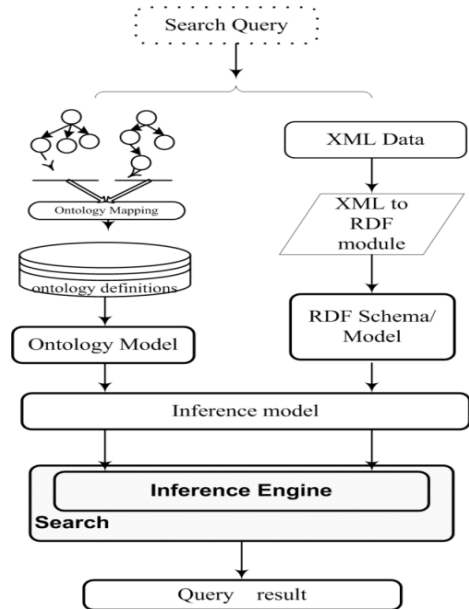


Fig.8. File retrieval in SemFARM.

6.1 Evaluating File Retrieval in SemFARM

A supplementary ontology was developed for evaluating the performance of SemFARM empowered by OARS. The domain concept of the supplementary ontology was selected from a sub-concept of the main generic ontology which was used in the implementation of SemFARM. The main purpose was to evaluate the effectiveness of OARS in support of ontology alignment in SemFARM.

6.2 Performance Evaluation Environments

Two case studies were considered for evaluation purposes which are given below:

- Case-1: SemFARM without OARS**
 One generic ontology was utilized in the setup to retrieve the required files. In this case, the search module of SemFARM utilized the knowledge extracted from the main ontology only. Hence, a single ontology was utilized in this setup therefore the alignments were not required and used.
- Case-2: SemFARM with OARS**
 Two ontologies were utilized in the setup to retrieve the required files. The search module of SemFARM utilized

OARS which aligns the main and second ontologies. In this case, more knowledge was obtained using two ontologies.

Three sets of tests were formulated in order to demonstrate the efficacy of ontology alignment of OARS in file retrieval on mobile devices. The final precision and recall values were calculated by an average of the three test results. In each test set, varied numbers of files were annotated with such keywords which were considered as relevant to the file-searching query. Varied numbers of relevant files were used in order to obtain the values of recall. It should be noted that some of the files were also annotated with such keywords which were not defined by the main ontology. However, these keywords were defined in the second ontology but with a limited concept domain.

Furthermore, varied query-words were used in each test set but it was made assured that the query-words contain the keywords defined in both ontologies to give a fair chance to both cases. Similarly, the same query-words were used for both cases in each corresponding test.

6.3 Computing Precision and Recall

An overall comparison of the two cases indicates an improvement of Case-2 over Case-1 in terms of precision against the same values of recall as shown in Fig.9.

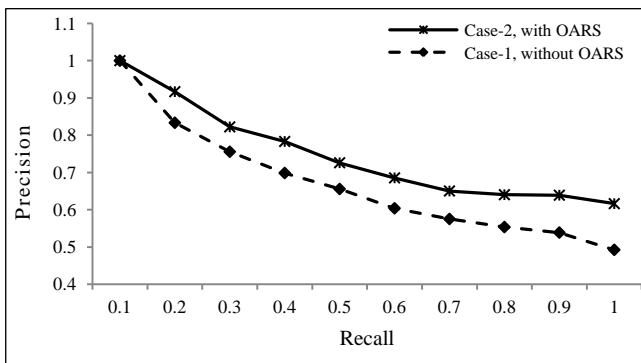


Fig.9. The performance of SemFARM empowered by OARS.

The average precision values of Case-1 and Case-2 are 0.65 and 0.72 respectively against the same recall value of 0.5. It can also be observed that the decrease in precision values in Case-2 is less than that of Case-1 as the recall value changes from 0.1 to 1. This can be further elucidated by the results showing that the precision values decrease from 1 to 0.49 in Case-1 and from 1 to 0.61 in Case-2 as the corresponding recall values increase from 0.1 to 1. The precision values for Case-2 and Case-1 are 0.616 and 0.492 respectively when recall value is 1. It is noted that the precision values are the same in both cases at the recall value of 0.1. The reason is that when the number of retrieved files is small, it is highly likely that the retrieved files would be relevant.

7. CONCLUSION

In this paper we have presented OARS, an ontology

alignment system using Rough sets to deal with uncertain entities in mapping. The use of Rough sets has proven to be effective in mapping entities for which the individual matchers cannot reach a decision in ontology mapping. The significance of using Rough sets as an aggregation method was evaluated and compared with a number of existing alignment systems using the benchmark ontology data sets of the OAEI 2010. The evaluation results are highly encouraging. The effectiveness of OARS was further evaluated in the SemFARM framework for enhanced file retrieval on mobile devices.

Currently, we are investigating the verification process of OARS in order to improve its performance in precision without degrading recall. For this purpose, we are planning to use the similarity of hierarchical path information between ontology entities. We also plan to participate in the OAEI campaign in the future. It is worth noting that OARS in its current form cannot align such ontologies where foreign languages are used to represent the class entities. Therefore, we are planning to integrate a few foreign (non-English) language dictionaries into the linguistic matcher to enable OARS to align ontologies defined in different language.

REFERENCES

- [1] D. Fensel, "Ontologies, Silver Bullet for Knowledge Management and Electronic Commerce", Springer-Verlag, 2001.
- [2] T. R. Gruber, "Toward Principles for the Design of Ontologies Used for Knowledge Sharing", *International Journal of Human Computer Studies*, vol. 43, no 5-6, pp. 907-928, 1995.
- [3] Y. Kalfoglou and M. Schorlemmer, "Ontology Mapping: The State of the Art", *The Knowledge Engineering Review*, vol. 18, no. 1, pp. 1-31, 2003.
- [4] P. Shvaiko, J. Euzenat, "A Survey of Schema-based Matching Approaches", *Journal on Data Semantics IV*, vol. 3730, pp. 146-171, 2005.
- [5] N. F. Natalya "Semantic Integration: A Survey of Ontology-based Approaches", *ACM SIGMOD Record*, vol. 33, no. 4, pp. 65-70, 2004.
- [6] E. Rahm, P. A. Bernstein, "A Survey of Approaches to Automatic Schema Matching", *The VLDB Journal, The International Journal on Very Large Data Bases*, vol.10, no. 4, pp. 334-350, December 2001.
- [7] J. Bock and J. Hettchenhausen, "MapPSO Results for OAEI 2008", in *Proc. of the 7th International Semantic Web Conference*, 2008.
- [8] J. Tang, J. Li, B. Liang, X. Huang, Y. Li and K. Wang, "Using Bayesian Decision for Ontology Mapping", *Journal of Web Semantics*, vol. 4, no. 1, pp. 243-262, 2006.
- [9] W. Hu, Y. Qu, "Falcon-AO: A Practical Ontology Matching System", *Journal of Web Semantics*, pp. 237-239, vol. 6, no. 3, 2008.
- [10] Z. Pawlak, "Rough Sets: Theoretical Aspects of Reasoning About Data", Kluwer Academic, Dordrecht, The Netherlands, 1991.
- [11] S. Jan, M. Li, G. Al-Sultany, H. Al-Raweshidy, "Ontology Alignment using Rough Sets", in *Proc. of the 8th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, pp. 2683-2686, 2011.
- [12] S. Jan, M. Li, G. Al-Sultany, Hamed Al-Raweshidy and I.A. Shah, "Semantic File Annotation and Retrieval on Mobile Devices", *Mobile Information Systems*, vol. 7, no. 2, pp. 107-122, 2011.
- [13] S. Jan, M. Li, G. Al-Sultany and H. Al-Raweshidy, "File Annotation and Sharing on Low-End Mobile Devices", in *Proc. of IEEE Interna-*

- tional Conference on Fuzzy Systems and Knowledge Discovery*, pp. 2973-2977, 2010.
- [14] M. Rodriguez and M. Egenhofer, "Determining Semantic Similarity among Entity Classes from Different Ontologies", *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, no. 7, pp. 442-456, 2003.
- [15] A. Tversky, "Features of Similarity", *Psychological Review*, vol. 84, no. 4, pp. 327-352, 1977.
- [16] J. Kwan and H. S. Yong, "Ontology Matching based on Hypemym, Hyponym, Holonym, and Meronym Sets in Word Net", *International Journal of Web & Semantic Technology*, vol. 1, no. 2, pp. 1-14, 2010.
- [17] M. Yatskevich and F. Giunchiglia, "Element Level Semantic Matching using WordNet", in *Proc. of International Conference on Meaning Coordination and Negotiation Workshop (ISWC)*, 2004.
- [18] W. Hu, N. S. Jian, Y. Z. Qu, and Y. B. Wang, "GMO: A Graph Matching for Ontologies", in *Proc. of K-Cap 2005 Workshop on Integrating Ontologies*, pp. 43-50, 2005.
- [19] Y. Qu, W. Hu, and G. Cheng, "Constructing Virtual Documents for Ontology Matching", in *Proc. of the 15th International World Wide Web Conference (WWW'06)*, pp. 23-31, 2006.
- [20] J. Li, J. Tang, Y. Li, and Q. Luo, "RiMOM: A Dynamic Multi-Strategy Ontology Aignment Framework", *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 8, pp. 1218-1232, 2009.
- [21] G. Stoilos, G. Stamou, S. Kollias, "A String Metric for Ontology Alignment", In *proc. of the 4th International Semantic Web Conference*, Springer LNCS, vol. 3729, pp. 624-637, 2005.
- [22] Y. R. Jean-Mary, E. P. Shironoshita, M. R. Kabuka, "Ontology Matching with Semantic Verification", *Journal of Web Semantics*, vol. 7, no.3, pp. 235-251, 2009.
- [23] P. Xu, Y. Wang, L. Cheng and T. Zang, "Alignment Results of SOBOM for OAEI 2010", in *Proc. of International Semantic Web Conference on Ontology Matching*, 2010.
- [24] I. F. Cruz, F. P. Antonelli, and C. Stroe, "AgreementMaker: Efficient Matching for Large Real-World Schemas and Ontologies", *PVLDB*, vol. 2, no. 2, pp. 1586-1589, 2009.
- [25] J. Noessner and M. Niepert, "CODI: Combinatorial Optimization for Data Integration—Results for OAEI 2010", in *Proc. of the 5th International Workshop on Ontology Matching*, 2010.
- [26] F. Hamdi, B. Safar, N. Niraula, and C. Reynaud, "TaxoMap in the OAEI 2009 Alignment Contest", in *Proc. of the ISWC Workshop on Ontology Matching*, pp. 230-237, 2009.
- [27] W. Hu, Y. Zhao, and Y. Qu., "Partition-based Block Matching of Large Class Hierarchies", in *Proc. of the 1st Asian Semantic Web Conference*, vol. 4185 of LNCS Springer-Verlag, pp. 72-83, 2006.
- [28] J. Bock and J. Hettenhausen, "MapSO Results for OAEI 2008", in *Proc. of the 7th International Semantic Web Conference*, 2008.
- [29] J. Bock, J. Hettenhausen, "Discrete Particle Swarm Optimisation for Ontology Alignment", *Information Sciences*, vol. 192, pp. 152-173, 2012.
- [30] H. Wache, T. Voegelé, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann, and S. Huebner, "Ontology-based Integration of Information - A Survey of Existing Approaches", in *Proc. of the workshop on Ontologies and Information Sharing at the International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 108-117, 2001.
- [31] O. Světlík and V. Svátek, "Combining Ontology Mapping Methods Using Bayesian Networks" in *Proc. of the International Semantic Web Conference*, 2006.
- [32] R. Pan, Z. Ding, Y. Yu, and Y. Peng, "A Bayesian Network Approach to Ontology Mapping", in *Proc. of International Semantic Web Conference*, pp. 563-577, 2005.
- [33] M. Nagy, M. Vargas-Vera, E. Motta, "DSSim-Ontology Mapping with Uncertainty", in *Proc. of 1st International Workshop on Ontology Matching (OM-2006)*, 2006.
- [34] H. Li, X. Zhou and B. Huang, "Method to Determine α in Rough set Model based on Connection Degree", *Journal of Systems Engineering and Electronics*, vol. 20, no. 1, pp.98-105, 2009.
- [35] M. Benerecetti, P. Bouquet and C. Ghidini, "Contextual Reasoning Distilled", *Journal of Experimental and Theoretical Artificial Intelligence*, vol. 12, no. 3, pp. 279-305, 2000.
- [36] C. Ghidini and F. Giunchiglia, "A Semantics for Abstraction", in *Proc. of the 16th European Conference on Artificial Intelligence*, pp. 343-347, 2004.
- [37] J. Euzenat and P. Shvaiko, "Ontology Matching", Springer, Heidelberg 2007.
- [38] J. Euzenat, et al., "State of the Art on Ontology Alignment", <http://disi.unitn.it/~p2p/RelatedWork/Matching/kweb-223.pdf> (Last accessed: 25 January 2012)
- [39] L. Levenstein, "Binary Codes Capable of Correcting Deletions, Insertions and Reversals", *Cybernetics and Control Theory*, pp. 707-710, 1966.
- [40] M. Jaro, "Probabilistic Linkage of Large Public Health Data Files", *Statistics in Medicine*, vol. 14, no. 5-7, pp. 491-498, 1995.
- [41] W. Winkler, "The State Record Linkage and Current Research Problems", *Technical Report*, Statistics of Income Division, Bureau of the Census, Washington, DC, 1999.
- [42] S. B. Needleman and C. D. Wunsch, "A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of two Proteins", *Journal of Molecular Biology*, vol. 48, pp. 443-453, 1970.
- [43] W. Cavnar and J. Trenkle, "N-Gram-Based Text Categorization", in *Proc. of Workshop on Document Analysis and Information Retrieval*, Las Vegas, pp. 161-169, Apr. 1994.
- [44] G. Kondrak, "N-Gram Similarity and Distance", in *Proc. of 20th Int. Conference on String Processing and Information Retrieval*, pp. 115-126, 2005.
- [45] W. W. Cohen, P. Ravikumar, and S. E. Fienberg, "A Comparison of String Distance Metrics for Name-Matching Tasks", in *Proc. of the Workshop on Information Integration on the Web*, pp. 73-78, 2003.
- [46] D. Lin, "An Information-Theoretic Definition of Similarity", in *Proc. of 15th International Conf. on Machine Learning*, Morgan Kaufmann, pp.296-304, 1998.
- [47] G. A. Miller, "Wordnet: A Lexical Database for English", *Communications of ACM*, vol. 38, no. 11, pp. 39-41, 1995.
- [48] A. Budanitsky and G. Hirst, "Evaluating WordNet-based Measures of Lexical Semantic Relatedness", *Computational Linguistics*, vol. 32, no.1, pp.13-47, 2006.
- [49] S. Greco, B. Matarazzo, R. Słowiński, "Rough Sets Theory for Multicriteria Decision Analysis", *European Journal of Operational Research*, vol. 129, no. 1, pp. 1-47, 2001.
- [50] Z. Pawlak, "Rough Sets", *International Journal of Information & Computer Sciences*, vol. 11, pp. 341-356. 1982.
- [51] J. Makhoul, F. Kubala, R. Schwartz and R. Weischedel, "Performance Measures for Information Extraction", in *Proc. of DARPA Broadcast News Workshop*, 1999.
- [52] A. Skowron, "Rough Sets in KDD - Plenary Talk", in *Proc. of the 16th World Computer Congress on Intelligent Information Processing (IIP'00)*, pp. 1-14, 2002.
- [53] Y. Yao, "A Note on Definability and Approximations", *LNCS Transactions on Rough Sets VII*, Springer, pp. 274-282, 2007.
- [54] J. Euzenat, "An API for Ontology Alignment", in *Proc. of 3rd Inter-*

- national Semantic Web Conference (ISWC)*, pp. 698-712, 2004.
- [55] J. Makhoul, F. Kubala, R. Schwartz and R. Weischedel, "Performance Measures for Information Extraction", in *Proc. of DARPA Broadcast News Workshop*, 1999.
- [56] C. J. van Rijsbergen, "Information Retrieval", Butterworths, London, 1979.
- [57] H. H. Do, S. Melnik, and E. Rahm, "Comparison of Schema Matching Evaluations", in *Proc. of the Workshop on Web, Web Services and Database Systems*, pp. 221-237, 2002.
- [58] J. Euzenat, A. Ferrara, C. Meilicke, J. Pane, F. Schar e, P. Shvaiko, H. Stuckenschmidt, O. Svab-Zamazal, V. Svatek, and C. Trojahn dos Santos, "Results of the Ontology Alignment Evaluation Initiative 2010", in *Proc. of the 5th ISWC Workshop on Ontology Matching (OM-2010)*, pp. 1-35, 2010.
- [59] S. Jan, M. Li and H. Al-Raweshidy, "File Annotation and Retrieval on Mobile Devices", *International Journal of Personal and Ubiquitous Computing*, Springer-Verlag London, vol. 15, no. 7, pp. 771-779, 2011.
- [60] Y. Wang, W. Liu and D. Bell, "Combining Uncertain Outputs from Multiple Ontology Matchers", In *proc. of the 1st International Conference on Scalable Uncertainty Management*, Lecture Notes in Computer Science, Springer, vol. 4772. pp. 201-214, 2007.
- [61] A. Isaac, S. Wang, C. Zinn, H. Matthezing, L. Meij, S. Schlobach, "Evaluating Thesaurus Alignments for Semantic Interoperability in the Library Domain", *IEEE Intelligent Systems*, vol. 24, no. 2, pp. 76-86, 2009.
- [62] A. Isaac, H. Matthezing, L. Meij, S. Schlobach, S. Wang, C. Zinn, "Putting Ontology Alignment in Context: Usage Scenarios, Deployment and Evaluation in a Library Case", in *Proc. of the 5th European Semantic Web Conference (ESWC)*, pp. 402-417, 2008.
- [63] S. Garruzzo and D. Rosaci, "Agent Clustering based on Semantic Negotiation", *ACM Transactions on Autonomous and Adaptive Systems*, vol. 3, no. 2, Article 7, 2008.
- [64] M. Sahami, T.D. Heilman, "A Web-based Kernel Function for Measuring the Similarity of Short Text Snippets", in *Proc. of 15th International World Wide Web Conference (WWW'06)*, pp. 377-386, 2006.
- [65] S. Banerjee, T. Pedersen, "Extended Gloss Overlaps as a Measure of Semantic Relatedness", in *Proc. of 18th International Joint Conference on Artificial Intelligence (IJCAI'03)*, pp. 805-810, 2003.
- [66] S. Patwardhan and T. Pedersen, "Using WordNet-based Context Vectors to Estimate the Semantic Relatedness of Concepts", in *Proc. of 11th Conference of European Chapter of the Association for Computational Linguistics (EACL'06), Workshop on Making Sense of Sense: Bringing Computational Linguistics and Psycholinguistics Together*, pp. 1-8, 2006.
- [67] D. Sánchez, A. SoléRibalta, M. Batet, and F. Serratos, "Enabling Semantic Similarity Estimation across Multiple Ontologies: An Evaluation in the Biomedical Domain", *Journal of Biomedical Informatics*, vol. 45, no. 1, pp. 141-155, 2012.
- [68] D. Sánchez, M. Batet, D. Isem, and A. Valls, "Ontology-based Semantic Similarity: A New Feature-based Approach", *Expert Systems with Applications*, vol. 39, no. 9, pp. 7718-7728, 2012.
- [69] N. Seco, T. Veale, J. Hayes, "An Intrinsic Information Content Metric for Semantic Similarity in WordNet", in *Proc. of the 16th European Conference on Artificial Intelligence (EACI'04)*, pp. 1089-1090, 2004.
- [70] D. Sánchez, M. Batet, D. Isem, "Ontology-based Information Content Computation", *Knowledge based Systems*, vol. 24, no. 2, pp. 297-303, 2011.

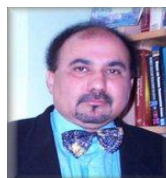


Dr. Sadaqat Jan is an Assistant Professor in Computer Software Engineering Department at Khyber Pakhtunkhwa University of Engineering and Technology, Mardan Campus, Pakistan. He received his PhD degree from Brunel University, UK and Master degree from NWFP University of Engineering and Technology Peshawar, Pakistan. His research interests are in the area of semantic web, mobile computing, information retrieval, grid computing, distributed systems, knowledge engineering and computer networks. He is a member of IEEE.



Dr. Maozhen Li is a Senior Lecturer in the School of Engineering and Design at Brunel University, UK. His research interests are in the areas of high performance computing (grid and cloud computing) for big data analysis, and intelligent systems. He is on

the Editorial Boards of Computing and Informatics journal and journal of Cloud Computing: Advances, Systems and Applications. He has over 100 research publications in these areas. He is a Fellow of the British Computer Society.



Professor Hamed Al-Raweshidy is currently the Director of the Wireless Networks and Communications Centre (WNCC) at Brunel University, UK. He is the editor of the first book in Radio over Fibre Technologies for Mobile Communications Networks and contributed to chapters for 6 books. Currently, he is the Editor-in-Chief of the Communication Networks Journal. He has acted as guest editor for the International Journal of Wireless Personal Communications. He is a member of several Journal Editorial Boards such as Journal of Communications and Mobile Computing and Journal of Wireless Personal Communications. He is a Fellow of the IET and a senior member of IEEE.

He is a member of several Journal Editorial Boards such as Journal of Communications and Mobile Computing and Journal of Wireless Personal Communications. He is a Fellow of the IET and a senior member of IEEE.



Dr. Alireza Mousavi is a Senior Lecturer in Systems Engineering and Computing. His research interest is in Smart Supervisory Control and Data acquisition Systems applied to real-time systems modelling and optimisation. The key areas of application are in stochastic modelling, ontology alignment and sensor networks. He is a member of both IEEE and IET.

He is a member of both IEEE and IET.



Dr. Man Qi is a Senior Lecturer in Department of Computing at Canterbury Christ Church University, UK. Her research interests are in the areas of computer graphics, computer animation, multimedia and applications. She is a Fellow of the British Computer Society and also a

Fellow of the Higher Education Academy.