

2008

A design and input-specific classification framework of ontology matching techniques

Kar Yin Leung
RMIT International University, Vietnam

S. Lau
University of Wollongong, simlau@uow.edu.au

Joshua P. Fan
University of Wollongong, joshua@uow.edu.au

Follow this and additional works at: <https://ro.uow.edu.au/gsbpapers>



Part of the [Business Commons](#)

Recommended Citation

Leung, Kar Yin; Lau, S.; and Fan, Joshua P.: A design and input-specific classification framework of ontology matching techniques 2008.
<https://ro.uow.edu.au/gsbpapers/287>

A design and input-specific classification framework of ontology matching techniques

Abstract

Ontology matching can be defined as the process of discovering similarities between two ontologies and it can be processed exploiting a number of different techniques. To provide a common conceptual basis, researchers have started to develop classifications to distinguish them. The most significant one is the classification proposed by Shvaiko and Euzenat to compare different existing ontology mediation systems as well as to design a new one. Although the classification is developed conscientiously, there are still some improper identifications and vague categories. Thus, a design and input-specific classification framework of ontology matching techniques that consists of executive approach, basic technique and input layers is proposed to address the above problems based on the findings of the literature survey. The framework provides not only a clear guideline on designing new mediation tool in accordance with the relationships among the three different layers but also an effective method to identify the type of the matching technique and its related executive approach simply by comparing input of mediation system with the input layer in the proposed framework.

Disciplines

Business

Publication Details

Leung, N. K. Y., Lau, S. Kim. & Fan, J. (2008). A design and input-specific classification framework of ontology matching techniques. In J. Richardson & A. Ellis (Eds.), Australian World Wide Web Conference (pp. 71-80). Lismore, NSW, Australia: Southern Cross University.

A Design and Input-specific Classification Framework of Ontology Matching Techniques

Nelson K. Y. Leung, Lecturer, Business Information Systems [HREF1], RMIT International University Vietnam [HREF2], Vietnam. nelson.leung@rmit.edu.vn

Sim Kim Lau [HREF3], Senior Lecturer, School of Information Systems and Technology [HREF4], University of Wollongong [HREF5], Australia. simlau@uow.edu.au

Joshua Fan [HREF6], Senior Lecturer, Graduate School of Business [HREF7], University of Wollongong [HREF8], Australia. joshua@uow.edu.au

Abstract

Ontology matching can be defined as the process of discovering similarities between two ontologies and it can be processed exploiting a number of different techniques. To provide a common conceptual basis, researchers have started to develop classifications to distinguish them. The most significant one is the classification proposed by Shvaiko and Euzenat to compare different existing ontology mediation systems as well as to design a new one. Although the classification is developed conscientiously, there are still some improper identifications and vague categories. Thus, a design and input-specific classification framework of ontology matching techniques that consists of executive approach, basic technique and input layers is proposed to address the above problems based on the findings of the literature survey. The framework provides not only a clear guideline on designing new mediation tool in accordance with the relationships among the three different layers but also an effective method to identify the type of the matching technique and its related executive approach simply by comparing input of mediation system with the input layer in the proposed framework.

1. Introduction

Ontology matching is one of the most important phrases in the process of ontology mapping and merging with the purpose of establishing semantic relationships between two ontologies. In general, ontology matching can be defined as the process of discovering similarities between two ontologies (Predoiu et al. 2006). It determines the relationships holding between two sets of entities that belong to two discrete ontologies (Shvaiko 2004). In other words, it is the process of finding a corresponding entity in the second ontology for each entity (for example, concept, relation, attribute and so on) in the first ontology that has the same or the closest intended meaning. This can be achieved by analysing the similarity of the entities in the compared ontologies in accordance with a particular metric (Ehrig and Sure 2004; INTEROP 2004). The correspondence can either be expressed by one to one functions or one to many functions. One to one functions denote that an entity in an ontology could have only one similar entity in another ontology whereas one to many functions address the fact that an entity may have more than one similar entities in another ontology (Castano et al. 2007).

This paper discusses the ontology matching process used in ontology mapping and merging. We investigate the classification and application of matching techniques as well as some of the most significant ontology mediation systems to develop a design and input-specific classification framework of ontology matching techniques. The aim of this framework is to provide a guideline for identifying the type of the matching technique and its related executive approach and for designing new mediation tool. This paper is organized as follows. Section 2 describes the background of ontology matching. Section 3 presents a literature review of ontology matching systems. A design and input-specific classification framework of ontology matching techniques is proposed in Section 4. Finally, conclusion is given at Section 5.

2. Background of Ontology Matching

Here, we use the mapping process developed by Ehrig and Staab (2004) to describe the five essential tasks of performing ontology matching (see Figure 1). To make the matching process more understandable, we slightly modify the tasks by categorizing them into three stages, namely pre-matching stage, matching stage and post-matching stage. In the pre-matching stage, some preparation works are required to complete before the actual similarity computation can take place. The pre-matching stage starts with feature engineering in which the initial representations of two ontologies are transformed into a common format suitable for similarity computation. In some cases, syntactic normalization is involved in the feature engineering task, that is, natural language processing techniques such as tokenization, lemmatization and elimination are adopted to normalize the syntactic heterogeneity

(Maedche et al. 2002; Giunchiglia et al. 2004). Another task in the pre-matching stage is to determine the next search step in finding matching candidate. The most common approach is to compare all entities of the first ontology with all entities of the second ontology (Noy and Musen 2000; McGuinness et al. 2000; Ehrig and Sure 2004). Other advanced approach allows the matching tool to choose computing the similarities of a subset of candidate concept pairs and to ignore others (Rahm et al. 2004). After the completion of the pre-matching stage, it comes with the matching stage where the actual similarity computation is carried out to determine the similarity values between matching candidates. Many researches are focused on developing mediation tools that adopt multiple matching techniques because it is unlikely for a single technique to achieve as many good matching candidates as the multiple one does (Rahm and Bernstein 2001). There are two ways to combine matching techniques, either by integrating multiple matching criteria in a hybrid matcher or by combining the results of independently executed matchers within a composite matcher (Di Martino 2006). As there may be more than one similarity values for a candidate concept pair, the post-matching stage requires the matching tool to aggregate different similarity values into a single value for one candidate pair. The final task of the post-matching stage requires the matching tool exploiting some sort of mechanisms to determine a suitable cut-off point. Thus, the cut-off point could interpret the similarity value in order to derive the best matching pair(s) among a concept in the first ontology and a set of concepts in the second ontology. These five tasks of ontology matching may need to iterate until no new similarities could be found.

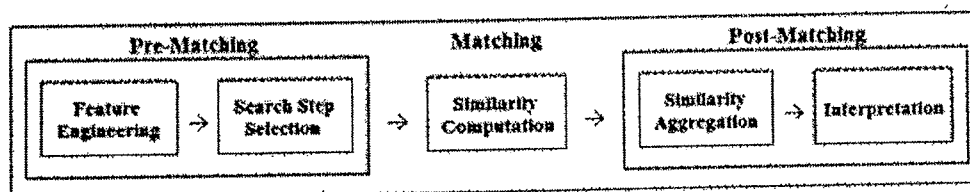


Figure 1: Ontology Matching Process (Ehrig & Staab 2004)

Ontology matching (or similarity computation) can be processed exploiting a number of different techniques. To provide a common conceptual basis, researchers have started to identify different types of ontology matching techniques and propose classifications to distinguish them, for example, Abels et al. (2005) propose a classification that consists of nine matching techniques based on existing literature studies. Another example is the classification developed by Shvaiko and Euzenat (2005). Building on the foundation of Rahm and Bernstein (2001)'s schema matching techniques classification, Shvaiko and Euzenat develop a meticulous classification to categorize elementary ontology and schema matching techniques. Their classification focuses on techniques that exploit ontology-level information excluding instance data. There are two synthetic classifications that can be viewed in top-down and bottom-up manner. The top-down view is called "granularity/input interpretation layer" which is based on granularity of match and then on how input information is interpreted. The bottom-up view is called "kind of input layer" and it is based on the kind of input requires in the matching process. "Granularity/input interpretation layer" and "kind of input layer" are further split into one common layer called "basic techniques layer". Ten different types of elementary matching techniques are identified in this layer:

- (1) String-based technique is used to match names and name descriptions of ontology entities in terms of a sequence of alphabet letters.
- (2) Language-based technique uses natural language processing techniques such as tokenization, lemmatization and elimination to exploit morphological properties of the input words. The technique is usually applied before string-based technique in order to improve the results.
- (3) Constraint-based technique is used to match the definitions of properties in terms of their internal constraints such as datatypes and cardinality.
- (4) Linguistic resources technique utilizes common knowledge or domain specific thesauri such as WordNet to analyse linguistic relations in the word matching process.
- (5) Alignment reuse technique exploits the idea of reusing alignments of previously matched ontologies as many ontologies to be matched are similar to already matched ontologies with the same application domain.
- (6) Upper level formal ontologies technique uses external source of common knowledge in the form of ontology such as DOLCE (Gangemi et al. 2003) within the matching process.
- (7) Graph-based technique considers the input as labelled graphs containing terms and their inter-relationships. Basically, the similarity is obtained through the analysis of the positions of a pair of nodes (from two ontologies) on the graphs.
- (8) Taxonomy-based technique also considers the input as graphs but the technique concerns only the specialization relation (is-a links).

- (9) Repository of structures technique stores ontologies and their fragments together with pair-wise similarities between them. When new ontologies are to be matched, the stored similarities could first be checked to avoid the matching operation to be performed over the dissimilar fragments. The available similarities could help to identify fragments that are worth carrying out matching in more detail.
- (10) Model-based technique deals with input based on its semantic interpretation using well grounded deductive methods such as propositional satisfiability and description logics.

3. A Literature Review of Ontology Matching Systems

In this section, we present a literature survey on some of the most significant mediation tools, frameworks and methods. Our focus is to examine their inherent matching process, in particular their similarity computation task at the matching stage of the process based on Shvaiko and Euzenat's classification. In this way, a detailed description could be provided to demonstrate how these ten matching techniques are performed in the actual mediation environment.

PROMPT framework is a multiple-ontology management tool that provides support to ontology versioning, ontology merging, ontology matching and other related management tasks. As one of the integrated components in the framework, iPROMPT assists users in the ontology merging process by suggesting matching candidates, identifying inconsistencies and potential problems as well as suggesting possible solution to resolve them (Noy and Musen 2000). iPROMPT uses simple heuristic approach to perform string-based ontology matching in the process of ontology merging. However, iPROMPT is only capable of matching classes with identical names from two different ontologies with no spelling deviation.

AnchorPrompt is another key component within the PROMPT framework that adopts heuristic approach to provide additional possible points of similarity between ontologies (Noy and Musen 2003). AnchorPrompt views each ontology as a directed labelled graph with node (class) and edge (slot) to represent the taxonomy of the ontology. To provide suggestions, AnchorPrompt first needs to identify, either manually or automatically, a set of pairs of related terms (anchors) from the source ontologies as input. With a pair of anchors, it is possible to define a number of paths that possess different sets of nodes and edges for each ontology. AnchorPrompt then traverses the paths in the corresponding ontologies and compare the nodes along to find similar terms. The similarity score is increased for the pairs of terms in the same position on the paths. This process is repeated for each pair of paths in order to generate the final score by aggregating the similarity score from all the traversals. In this way, AnchorPrompt is able to produce a set of semantically related concepts from the source ontologies.

Ontology Mapping Framework (MAFRA) is a mapping framework that provides a generic view onto the overall mapping process for distributed ontologies in the semantic web (Maedche et al. 2002). MAFRA uses a relatively more complex heuristic approach to perform ontology matching. On one hand, MAFRA makes use of two external linguistic resources to find lexical similarities between two ontologies in terms of concepts, attributes and relations. On the other hand, MAFRA exploits constraint-based technique to acquire similarity between concepts based on their properties. Furthermore, taxonomy-based matching technique is applied to the bottom-up and top-down strategy. The bottom-up strategy takes property similarity as input to propagate from lower parts of the ontology to the upper concepts. Simultaneously, the top-down strategy also allows MAFRA to propagate the similarity from top to bottom.

Naïve Ontology Mapping (NOM) is an ontology mapping approach that integrates various similarity measuring methods to identify possible mappings (Ehrig and Sure 2004). At present, seventeen manually encoded matching rules are used to measure similarities of two ontologies in the aspects of concepts, relations, instances and files. Other than the two rules designed for measuring similarities of files, the remaining fifteen rules are developed based on four types of matching techniques. First of all, string-based technique is adopted to compare labels or URI of concepts, relations and instances between two different ontologies. Second, linguistic resources technique is used for carrying out comparisons across languages. Third, constraint-based technique is applied to match properties between two ontologies. Fourth, taxonomy-based technique is used to acquire similarity derived from the super(sub)-concepts and super(sub)-properties relationships. The similarity results of the seventeen rules are aggregated and interpreted using different combinations of aggregation algorithms and cut off point determination strategies.

Quick Ontology Mapping (QOM) builds upon the success of NOM and is grounded on its seventeen matching rules (Ehrig and Stabb 2004). Instead of focusing solely on the quality of mapping results, QOM ameliorates the efficiency of the process. To do so, QOM uses heuristic to discard the less promising mapping candidates in order to lower the number of candidates to be compared in the matching process. In addition, the use of some of the costly features in the rules has been restricted so as to optimize the matching efficiency.

Glue is a semi-automatic system which applies machine learning to perform taxonomy-based ontology matching with the purpose of creating semantic mappings between two ontologies (Doan et al. 2002). Thus, the focus of the system is on finding the most similar concept in the second ontology for each concept in the first ontology. To do so, Glue first takes two taxonomies (for example O1 and O2) of two ontologies and their instances as input. Given that A and B are concepts of O1 and O2 respectively, machine learning is applied to allow a set of base learners to classify whether every instance of O1 is also an instance of concept A and whether every instance of O2 is also an instance of concept B. Meta-learner then combines the classifications from multiple base learners. The combined classifications are used to compute the joint probability distribution that consists of four probabilities: $P(A,B)$, $P(A,\neg B)$, $P(\neg A,B)$, and $P(\neg A,\neg B)$. Subsequently, Glue employs a user-supplied similarity algorithm to compute a similarity value for each pair of concepts in form of a matrix. Finally, Glue applies relaxation labeling technique to search for an appropriate set of mapping configurations based on similarity values in the matrix, domain constraints and common knowledge.

IF-Map is an automatic mapping method that uses model-based matching technique in its mapping process (Kalfoglou and Schorlemmer 2003). The technique is grounded on Channel Theory which provides a mathematical model to depict the flow of information in the connection channel between communities by means of tokens and types (Barwise and Seligman 1997). Thus, IF-Map associates ontology with local logics (a set of concepts, instances and relations) and formalizes the mappings in terms of logic informorphisms (morphisms between local logics). Other than using model-based matching technique, IF-Map as well exploits a set of heuristics as well as string-based and taxonomy-based techniques to derive partial mappings between the concepts of the ontologies.

Combining Match Algorithm (COMA) is an ontology matching system that provides a platform to combine multiple matchers in a flexible way (Do and Rahm 2002). Its flexibility not only allows user to choose from a wide variety of single matchers such as string-based, constraint-based and linguistic resources matchers, it also provides a number of hybrid matchers where different match criteria or properties are implemented in one fixed algorithm to serve in a specific way, for example, name matcher combines several string-based, language-based and linguistic resources matchers in order to derive similarities between element names. Apart from that, COMA offers an innovative reuse-oriented matcher in which user can reuse previous match results or alignments for entire new ontologies or for its fragments. Furthermore, the above simple matchers and hybrid matchers can be combined to form a composite matcher in accordance with the nature of the matching task at hand. COMA divides matching process into three phases: an optional user feedback phase, the execution of various matchers as well as the aggregation and interpretation of similarity results. User can specify the matching process to take place in one or multiple iterations by selecting from either the interactive mode or automatic mode. In the interactive mode, user has to, for each iteration, determine matchers' combination, similarity aggregation algorithm and a proper cut off point, define match or mismatch relationships as well as accept or reject matching candidates in the previous iteration. In the automatic mode, the matching process iterates only once with a default matching strategy (default matcher(s), aggregation algorithm and cut off point).

OWL-Lite Alignment (OLA) is designed to provide an environment for manipulate alignments of ontologies expressed in OWL, with an emphasis on OWL-Lite (Euzenat et al. 2004). In OLA, entities are compared according to their categories (class, object, property, relation, property instance, datatype, data value and property restriction label) using the same similarity function and on the same feature space. OLA uses a labelled graph to describe the ontology in which nodes and edges are used to represent entities and relationships respectively. Similarity measure of OLA is then defined by a system of quasi-linear equations and its similarity values are derived by means of an iterative approximation process which starts with measuring labels of the nodes and gradually expands to their neighbouring nodes. The similarity measuring model used to compare two nodes of different ontologies depends on the similarity of their labels, neighbouring nodes and other descriptive knowledge. Hence, OLA adopts string-based, language-based, linguistic resources matching techniques to execute the above measuring model.

S-Match is an automatic matching system that takes two graph-like taxonomies as input and computes the strongest semantic relations holding between any pair of nodes from two taxonomies (Giunchiglia and Shvaiko 2003; Giunchiglia et al. 2004). The match approach of S-Match is based on two notions, the notion of concept of a label and the notion of concept of a node. While the former one is defined as the set of documents that reflects the meaning of the labels, the latter is defined as the set of documents classified under this node. Before executing the semantic matching process, the system uses natural language processing techniques to tokenize and lemmatize labels of the taxonomies. Subsequently, the labels are required to translate from natural language into more precise internal format using WordNet. For complex concepts, logical connectives are used to replace tokens that contain prepositions, punctuation marks and conjunctions. To derive the strongest relationships, S-Match has to compute the concept of label matrix which contains relations between any two concepts of labels in the two

taxonomies using both the string-based and linguistic resources matchers. After that, the information obtained from the above matrix together with concepts of labels and concepts of nodes are codified into a set of complex propositional formulas. Finally, S-Match could generate another matrix containing the strongest relations holding between concepts of nodes by applying model-based technique to process the propositional formulas.

A summary of the survey that consists of ten mediation tools, frameworks and methods with their inherent matching techniques is illustrated in Table 1. The most popular ontology matching techniques are string-based, taxonomy-based and linguistic resources techniques. Each of them is used by at least six out of the ten mediation systems. In contrast, the least popular matching technique is upper level formal ontologies and it is not adopted by any system at all. All systems in the survey incorporate a graph algorithm as their matching technique (either graph-based or taxonomy-based technique). Most of the mediation systems exploit multiple matching strategies which exploit more than one matching techniques, for instance, COMA includes six matching techniques in their inherent matching strategy. Thus leaving iPROMPT and Glue to engage with a single strategy in which only one matching technique is included in each system. In terms of execution approach, heuristic is widely implemented for carrying out string-based, language-based, constraint-based, linguistic resources, alignment reuse, graph-based, taxonomy-based and repository of structures matching techniques. Probabilistic reasoning approach (such as machine learning) also plays a part in the execution of taxonomy-based technique whereas semantic reasoning is the dedicated approach used to execute model-based technique. Out of the ten mediation systems, six of them are capable of performing ontology matching automatically, three of them still rely on human intervention and the remaining one allows user to execute ontology matching either automatically or semi-automatically.

Table 1: Summary of Mediation Systems and their Inherent Matching Techniques

Systems	Automation	String-based	Language-based	Constraint-based	Linguistic Resources	Alignment Reuse	Upper Level Formal Ontologies	Graph-based	Taxonomy-based	Repository of Structures	Model-based
COMA	Full	Heuristic		Heuristic	Heuristic				Heuristic	Heuristic	
NOIM	Full	Heuristic		Heuristic	Heuristic				Heuristic		
S-Match	Full	Heuristic	Heuristic		Heuristic				Heuristic		Semantic Reasoning
IF-Map	Full	Heuristic							Heuristic		Semantic Reasoning
iPROMPT	Semi	Heuristic									
ANCHORPROMPT	Semi	Heuristic							Heuristic		
COMA	Full / Semi	Heuristic	Heuristic	Heuristic	Heuristic	Heuristic		Heuristic			
MAFRA	Full			Heuristic	Heuristic				Heuristic		
GLUE	Semi								Probabilistic Reasoning		
OLA	Full	Heuristic	Heuristic	Heuristic	Heuristic			Heuristic			

4. Critical Analysis of Classification of Ontology Matching Techniques

Shvaiko and Euzenat's classification is designed to standardize a conceptual basis for comparing different existing ontology mediation systems as well as for designing a new one. Before comparison can be taken place, it is necessary to identify the types of ontology matching techniques implemented in the mediation tools. One way to identify the matching technique is by examining its input, for example, the adoption of a taxonomy-based matcher is substantiated when a mediation tool takes two taxonomies that contain subset and superset relationships as inputs. However, the two synthetic views of the classification only provide a very general description on classifying the inputs which may become an obstacle for carrying out the identification task effectively. As the categories used to classify inputs are too vague, it is essential to replace them with a more specific input layer in the classification. In the aspect of designing new mediation tools, a more specific input layer is essential to illustrate the relationship between the ten elementary ontology matchers and their inputs. Apart from that, this classification lacks the depiction of executive approach used to perform the related ontology matching technique, for instance, semantic reasoning approach is always used to execute model-based technique (Giunchiglia and Shvaiko 2003; Kalfoglou and Schorlemmer 2003). It is necessary for mediation tool designers and developers to understand the relationship between the executive approaches and elementary matching techniques. Such an understanding not only provides a guideline for designing new tools but it also helps to speed up the duration of design and development. Thus, an additional layer

of executive approaches should be included in the classification.

Even though the ten elementary matching techniques are categorized conscientiously, there are still several improper identifications in between. The first improper identification is the language-based matching technique. As mentioned earlier, this technique is normally performed prior to string-based technique and has no direct engagement in the actual similarity computation between two ontologies. In fact, this technique is used to normalize the syntactic heterogeneity within the matching process. It is more appropriate to consider this natural language processing technique as the first task (feature engineering) at the pre-matching stage of the ontology matching process, rather than labelling it as one of the matching techniques. The second improper identification is the repository of structures technique. This technique is a dynamic approach used to compare fragments of two ontologies and eliminate the dissimilar portions with the purpose of improving computational efficiency and cost in the matching process. Rather than classifying as one of the matching techniques, it should be regarded as the second task (search step selection) at the pre-matching stage of the matching process. The third improper identification is upper level formal ontologies technique. Shvaiko and Euzenat (2005) state that there is currently no mediation system using this technique. Their finding is further confirmed by our research conducted in the previous section to review some of the most significant mediation systems. Because there is insufficient evidence to specify the input and design guideline of this technique, it is therefore not reasonable to include a non-existing technique in the classification.

4.1. A Design and Input-specific Classification Framework of Ontology Matching Techniques

The classification defined by Shvaiko and Euzenat comprehensively itemizes ten elementary matching techniques, but three of them are improperly identified. Furthermore, the lack of an executive approach layer and a detailed input layer magnifies its incapability of allowing researchers, scholars as well as tool designers and developers to perform the technique identification and tool designing tasks. Therefore, we propose to develop a design and input-specific classification framework of ontology matching techniques that consists of the above two layers. As shown in Figure 2, there are three main layers in the proposed framework, namely executive approach, basic technique and input layer. Language-based, upper level formal ontologies and repository of structures matching techniques are excluded from the proposed framework to address the misidentification problem, thus leaving only string-based, linguistic resources, constraint-based, alignment reuse, graph-based, taxonomy-based and model-based matching technique in the basic technique layer of the proposed framework.

There are two different ways to study the proposed framework either by middle view or bottom-up view. The middle view describes the relationships among elementary matching techniques, executive approaches and input types. This view not only indicates the approach required to execute a particular matching technique, for example, heuristic approach can be used to execute string-based technique, it also provides a guideline for designing new mediation tool, for instance, to exploit model-based as a mediation system's matching technique, tool designer must ensure the input type and executive approach are propositional formulas and semantic reasoning respectively. The bottom-up view provides an easier way to identify the type of ontology matching technique and its executive approach simply by comparing input of mediation system with the input types on the input layer, for example, the matcher is most likely string-based if it takes names and descriptions of entities as input.

In the executive approach layer, we identify heuristic, probabilistic reasoning and semantic reasoning as three major approaches to execute the above seven elementary ontology matching techniques. Heuristic approach exploits rules for comparing syntactic features, properties, linguistic and structural information of two or more different ontologies (Castano et al. 2007). This approach is widely used in the execution of string-based, constraint-based, linguistic resources, alignment reuse, graph-based and taxonomy-based matching technique. In terms of string-based matching technique, heuristic approach establishes rules to determine the matching entities based on the similarity computation of representational strings from two ontologies. Examples of string-based rules include:

- Two entities are identical if their representational strings are identical (Ehrig and Stabb 2004; Ehrig and Sure 2004; Kalfoglou and Schorlemmer 2003; McGuinness et al. 2000; Noy and Musen 2000; Noy and Musen 2003).
- Two entities are identical if their representational strings contain the same prefix or suffix (Aumueller et al. 2005; Do and Rahm 2002).
- The similarity of two entities is higher if the number of steps required to convert one representational string to another are lower (Aumueller et al. 2005; Do and Rahm 2002).

While the string-based matching technique focuses only on calculating the string similarity of properties between two ontologies, the constraint similarity of the properties are taken care by the constraint-based

technique. Here, heuristic approach applies rules to find matching properties based on the internal constraints that apply to each property. Examples of constraint-based rules are:

- The similarity of two properties is higher if the ranges of their datatypes are closer (Ehrig and Stabb 2004; Ehrig and Sure 2004; Euzenat et al. 2004).
- The similarity of two properties is higher if the cardinalities of their values are closer (Ehrig and Stabb 2004; Ehrig and Sure 2004; Euzenat et al. 2004).

Linguistic resources matching technique uses a common knowledge or a domain specific thesaurus to derive meanings of entities in ontologies. By taking these meanings as input, heuristic rule is capable of determining the linguistic relations (such as synonyms, hyponyms and hypernyms) among the entities, for example, if a linguistic resources matcher derives from a common knowledge thesaurus that "Laptop" in Ontology A is a hyponym of "Computer" of Ontology B, heuristic could determine "Laptop" in A is subsumed by "Computer" in B (Aumueller et al. 2005; Do and Rahm 2002; Euzenat et al. 2004).

Alignment reuse matching technique makes use of previously matching results at the level of ontology fragments or entire ontologies to derive new matching results. Heuristic of the technique is built on a transitive nature of the similarity relation between elements (Aumueller et al. 2005; Do and Rahm 2002). This transitive nature means that if x is similar to y and y is similar to z, then x is very likely similar to z. In other words, it allows heuristic to reuse the available alignment information for matching analysis when Ontology B and C are required to match with each other, given that the matching results between A and B as well as between A and C have been stored.

Graph-based matching technique takes two ontologies in the form of labelled graphs as input and from which nodes from the ontologies are compared and analysed to derive the similarity of their neighbouring nodes. Examples of rules adopted by heuristic of graph-based technique include:

- Two nodes are similar if their immediate children nodes are highly similar (Aumueller et al. 2005; Do and Rahm 2002).
- Two nodes are similar if their leaf nodes are highly similar (Aumueller et al. 2005; Do and Rahm 2002).
- Two nodes are similar if their relations are similar (Euzenat et al. 2004).

Similar to graph-based technique, taxonomy-based matching technique also takes graph as input. However, the graph intake here is more rigorous because neighbouring nodes on the graph are connected with is-a links to indicate they are superset/subset of each other. Heuristic can be applied to compare and identify similar nodes along the paths connected by is-a links (Ehrig and Stabb 2004; Ehrig and Sure 2004; Kalfoglou and Schorlemmer 2003; Noy and Musen 2000; Noy and Musen 2003). Examples of rules adopted by heuristic of taxonomy-based technique are:

- Two nodes are similar if their super-nodes are the same (Ehrig and Stabb 2004; Ehrig and Sure 2004).
- Two nodes are similar if their sub-nodes are the same (Ehrig and Stabb 2004; Ehrig and Sure 2004).

Alternatively, probabilistic reasoning approach, such as Bayesian network and machine learning, can also be used to execute the taxonomy-based technique (Doan et al. 2002; Mitra et al. 2005; Prasad et al. 2002). Probabilistic reasoning uses probability measurement to represent the similarity of two concepts from two different taxonomies that are similar or having the same instances (Castano et al. 2007). When two independent taxonomies contain a pair of similar nodes, for example Node A and B, it is possible to induce new set(s) of similar nodes from the taxonomies by considering the probabilistic similarity measured between Node A and the neighbours of Node B and between Node B and the neighbours of Node A.

To execute model-based matching technique, semantic reasoning approach first requires to translate relationships of all possible matching candidates of two ontologies into some forms of propositional formula, such as axioms, local logics and so on (Giunchiglia and Shvaiko 2003; Giunchiglia et al. 2004; Kalfoglou and Schorlemmer 2003). Subsequently, the approach adopts sound deduction methods to validate matching between two ontologies in accordance with the semantic of propositional formulas, for instance, propositional satisfiability solver is used to check possible matching candidates by validating their propositional formulas (Castano et al. 2007).

In the input layer, inputs are further classified into two levels. The first level of this layer contains two keywords used to sum up the characteristics of the actual inputs on the second level: elementary and structural. Elementary input represents input that undergoes analysis in isolation during the matching process without the need of considering its relations with other entities. Names and descriptions of entities as well as datatypes and values of properties are classified as elementary input.

- Name of an entity refers to a sequence of characters that are used to name any entity (such as label, relation and property) in an ontology.
- Description of an entity is defined as a sequence of characters that are used to describe any entity (such as label, relation and property) in an ontology.
- Datatype of a property refers to the type of data associated with a property.
- Value of a property is the numerical quantity assigned to a property.

In contrast, mediation systems analyse structural input in accordance with its relations with other entities in the process of ontology matching. Alignments, graphs, taxonomies and propositional formulas are categorized as structural input.

- Alignment is the matching result obtained from previously matched ontologies.
- Graph is a type of structural representation that contains nodes and their inter-relationships.
- Taxonomy is similar to a graph except its nodes are connected with each other by is-a links, that is, the connected node is either subset or superset of another.
- Propositional formula refers to a semantic statement that describes a pair of possible matching candidates and their relations.

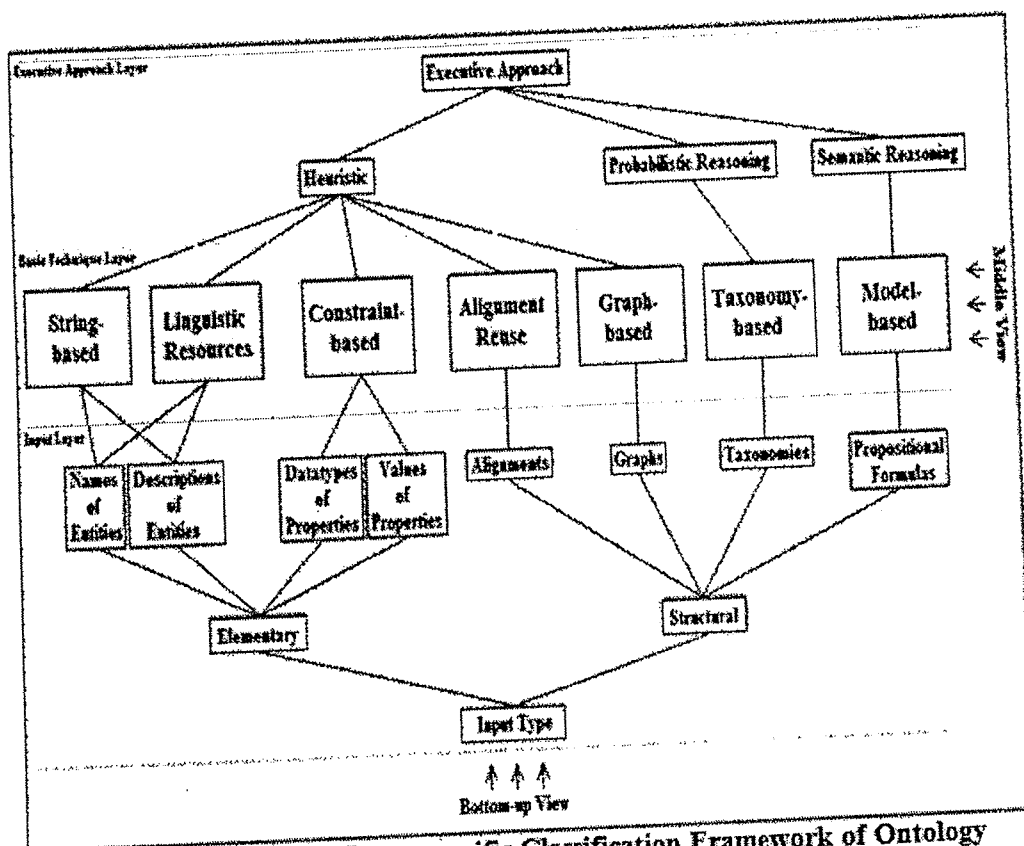


Figure 2: A Design and Input-specific Classification Framework of Ontology Matching Techniques

5. Conclusion

Shvaiko and Euzenat's classification aims to provide a conceptual basis of ontology matching techniques for comparing existing ontology mediation system and for designing a new one. Unfortunately, the improper identification of matching techniques as well as the lack of an executive approach layer and a detailed input layer makes it impossible to do so. As a result, we propose a design and input-specific classification framework of ontology matching techniques to address the above problems based on the findings of the literature survey. The proposed framework consists of three layers, namely executive approach, basic technique and input layer. On one hand, the proposed framework provides a clear guideline on designing new mediation tool based on the middle view that describes the relationships among the three different layers. On the other hand, the bottom-up view provides an effective method to identify the type of the matching technique and its related executive approach simply by comparing input of mediation system with the input layer in the proposed framework.

References

- Abels, S., Haak, L. and Hahn, A. (2005) "Identification of Common Methods Used for Ontology Integration Tasks" in *Proceedings of the 1st International ACM Workshop on Interoperability of Heterogeneous Information Systems (IHIS)* p.75-78.
- Aumüller, D., Do, H., Massmann, S. and Rahm, E. (2005) "Schema and Ontology Matching with COMA++" in *Proceedings of the ACM SIGMOD International Conference on Management of Data* p.906 – 908.
- Barwise, J. and Seligman, J. (1997) *Information Flow: The Logic of Distributed Systems*. Cambridge University Press.
- Castano, S., Ferrara, A., Montanelli, S., Hess, G. N. and Bruno, S. (2007) *State of the Art on Ontology Coordination and Matching*. BOEMIE Bootstrapping Ontology Evolution with Multimedia Information Extraction Project, FP6-027538 D4.4.
- Di Martino, B. (2006) "An Ontology Matching Approach to Semantic Web Services Discovery" in *Lecture Notes in Computer Science (LNCS)* v.4331 p.550–558.
- Do, H. and Rahm, E. (2002) "COMA – A System for the Flexible Combination of Schema Matching Approaches" in *Proceedings of the 28th International Conference on Very Large Databases (VLDB)* p.610-621.
- Doan, A., Madhavan, J., Domingos, P. and Halevy, A. (2002) "Learning to Map between Ontologies on the Semantic Web" in *Proceedings of the 11th International Conference on World Wide Web* p.662-673.
- Ehrig, M. and Stabb, S. (2004) "QOM – Quick Ontology Mapping" in *Lecture Notes in Computer Science (LNCS)*, v.3298 p.683–697.
- Ehrig, M. and Sure, Y. (2004) "Ontology Mapping – An Integrated Approach" in *Lecture Notes in Computer Science (LNCS)*, v.3053 p.76–91.
- Euzenat, J., Loup, D., Touzani, M. and Valtchev, P. (2004) "Ontology Alignment with OLA" in *Proceedings of the 3rd International Workshop on the Evaluation of Ontology-based Tools (EON)* p.333-337.
- Gangemi, A., Guarino, N., Masolo, C. and Oltramari, A. (2003) "Sweetening WordNet with DOLCE" in *AI Magazine* v.24 n.3 p.13-24.
- Giunchiglia, F. and Shvaiko, P. (2003) "Semantic Matching" in *The Knowledge Engineering Review Journal* v.18 n.3 p.265-280.
- Giunchiglia, F., Shvaiko, P. and Yatskevich, M. (2004) "S-Match: an Algorithm and an Implementation of Semantic Matching" in *Lecture Notes in Computer Science (LNCS)* v.3053 p.61-75.
- INTEROP (2004) *Ontology Interoperability*. State of the Art Report (SOA), WP8ST3 Deliverable, IST-508011.
- Kalfoglou, Y. and Schorlemmer, M. (2003) "IF-Map: An Ontology-Mapping Method based on Information-Flow Theory" in *Journal on Data Semantics 1* p.98-127.
- Maedche, A., Motik, B., Silva, N. and Volz, R. (2002) "MAFRA – A Mapping Framework for Distributed Ontologies" in *Lecture Notes in Computer Science (LNCS)* v.2473 p.235–250.
- McGuinness, D., Fikes, R., Rice, J. and Widler, S. (2000) "An Environment for Merging and Testing Large Ontologies" in *KR2000: Principles of Knowledge Representation and Reasoning* p.483-493.
- Mitra, P., Noy, N. and Jaiswal, A. (2005) "Ontology Mapping Discovery with Uncertainty" in *Lecture Notes in Computer Science (LNCS)* v.3729 p.537-547.
- Noy, N. and Musen, M. (2000) "PROMPT: Algorithm and tool for automated ontology merging and alignment" in *Proceedings of the 17th National Conference on Artificial Intelligence (AAAI)*.
- Noy, N. and Musen, M. (2003) "The PROMPT Suite: Interactive Tools for Ontology Merging and Mapping" in *International Journal of Human-Computer Studies* v.59 n.6 p.983-1024.
- Prasad, S., Peng, Y. and Finin, T. (2002) "Using Explicit Information to Map between Two Ontologies" in

Proceedings of the 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS).

Predoiu, L., Feier, C., Scharffe, F., de Bruijn, J., Martin-Recuerda, F., Manov, D. and Ehrig, M. (2006) *State-of-the-art Survey on Ontology Merging and Aligning V2*. EU-IST Integrated Project (IP) IST-2003-506826 SEKT: Semantically Enabled Knowledge Technologies, University of Innsbruck.

Rahm, E. and Bernstein, P. (2001) "A Survey of Approaches to Automatic Schema Matching" in *The International Journal on Very Large Data Bases (VLDB)* v.10 n.1 p.334-350.

Rahm, E., Do, H. and MaBmann, S. (2004) "Matching Large XML Schemas" in *SIGMOD Record* v.33 n.4 p.26-31.

Shvaiko, P. (2004) "A Classification of Schema-based Matching Approaches", In *Proceedings of the Meaning Coordination and Negotiation Workshop (MCN) at the 1st International Semantic Web Conference (ISWC)*".

Shvaiko, P. and Euzenat, J. (2005) "A Survey of Schema-based Matching Approaches" in *Journal on Data Semantics IV* p.146--171.

Hypertext References

HREF1 <http://www.rmit.edu.vn/undergraduate.php?uid=4>

HREF2 <http://www.rmit.edu.vn>

HREF3 <http://www.uow.edu.au/~simlau>

HREF4 <http://www.uow.edu.au/informatics/sisat/>

HREF5 <http://www.uow.edu.au>

HREF6 <http://www.uow.edu.au/gsb/staff/UOW025586.html>

HREF7 <http://www.uow.edu.au/gsb/>

HREF8 <http://www.uow.edu.au>