

A Survey on Ontology Matching Techniques

Satish Kumar
National Institute of Technology,
Kurukshetra, India.
er.satishkumar149@gmail.com

Vikram Singh
National Institute of Technology,
Kurukshetra, India.
viksnitkr@outlook.com

Balwinder Saini
National Institute of Technology,
Kurukshetra, India.
me7saini@gmail.com

Abstract— Words, sentences, paragraphs and documents is an important component in the data mining application area such as information retrieval, text summarization etc. And finding the similarity value between these components is an important action. Now a day, there are various matching techniques are present which are very helpful for finding the similarity between the ontology. In this survey paper discusses the existing works on similarity (On text, words etc.) using element-level techniques and structure level techniques.

Index Terms— Similarity, String-Based similarity, Element-Level techniques, Structure-Level techniques

I. INTRODUCTION

Homogeneous attribute measures play a consequential role in text cognate area and application in tasks such as ontology matching, information retrieval, query answering etc [1, 2]. Measuring the kindred attribute between the words, text, documents is the primary stage in ontology matching. Ontology matching is a paramount operation traditional applications e.g., Ontology evolution, Ontology integration, data integration and data warehouses [2]. In text similarity words can be similar in two ways lexically and semantically. If the sequence of character is similar then they are lexically similar and if the two words are opposite of each other having similar meaning then they are semantically similar [3]. These types of similarities measures through the element-level techniques between two ontologies. Element -Level techniques discusses the String-predicated techniques, Language-predicated techniques and Constraint-Based techniques etc.

On the other hand Structure-Level techniques are used to measure the similarity based on the structure. Structure-Level techniques discuss the Graph-predicated technique, Taxonomy-predicated technique and Constraint-predicated technique.

II. MATCHING PROCESS

The ontology matching process is utilized to measures the homogeneous attribute between the two ontologies [4]. The matching process determines an alignment A' , for two ontologies O_1 and O_2 . There are some other parameters of the matching process.

(i) The utilization of an input alignment A

(ii) Matching parameter like weights, threshold

(iii) External resources utilized by the matching process like erudition-base and domain-categorical area.

Definition 1 (Matching process) The matching process considered as function f which emanates from the ontologies to match O_1 and O_2 , an input alignment A , a set of parameters p and a set of oracles and resources r , back to an alignment A' between these ontologies[1, 4].

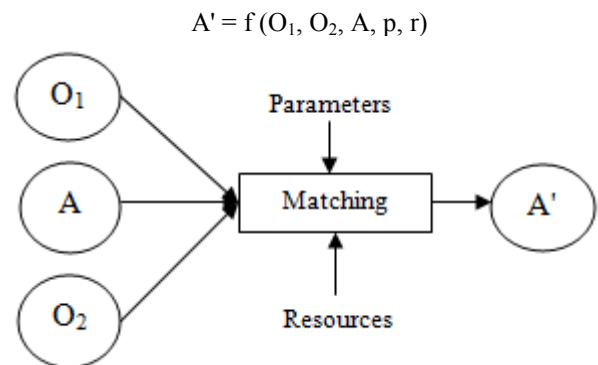


Fig. 1 The Matching Process

III. TYPES OF ONTOLOGY MATCHING TECHNIQUES

There are two types of otology matching techniques discusses. First one is Element-level technique and another one is Structure-level techniques [4].

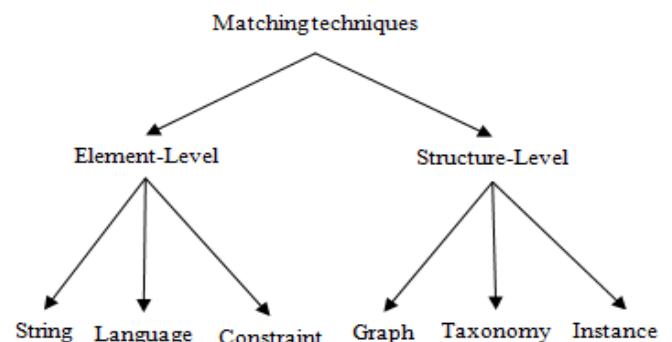


Fig. 2 Types of Ontology Matching Techniques

A. Element-Level technique

Element-Level matching techniques measure the homogeneous attribute of ontology and their instances in isolation from their relation with the other attribute and their instances [4]. Element-Level matching techniques discuss the String-predicated, Language-predicated and constraint-predicated methods.

String-predicated technique

A string-predicated technique is utilized to match the designations and name description of ontology entities. This technique takes string as an input and considers the strings as a sequence of letter in an alphabet [4]. In this technique various distance measures methods are utilized for finding the distance between strings such as Edit-distance, Hamming-distance etc. String-predicated method considers the structure of the string (as a sequence of letters) [5]. There are many ways to find the homogeneous attribute between the string; on the way the string is viewed as: as an exact sequence of letters, an erroneous sequence of letters, a set of letters, and a set of words. In this paper most frequently methods are discussed.

String parity

String parity defines the homogeneous attribute between the strings. If the strings are identical the string parity returns 1 and if the strings are not identical then it returns 0.

S = set of strings

s, t = are the strings

$|s|$ = length of the string s

$|t|$ = length of the string t

$s[i]$ for $i \in [1, |s|]$ latter at position i of s .

Definition 2 (String Parity) String parity is similarity $\sigma : S * S \rightarrow [0, 1]$ such that $\forall x, y \in S, \sigma(x, x) = 1$ and if $x \neq y, \sigma(x, y) = 0$.

This method does not explain the how the strings are dissimilar. Another way is also used for comparison two strings that is Hamming distance which shows and counts the number of positions where the strings are dissimilar.

Definition 3 (Hamming Distance) The Hamming distance [4, 6] is dissimilarity $\delta : S * S \rightarrow [0, 1]$ such that

$$\delta(s, t) = \frac{\left(\sum_{i=1}^{\min(|s|, |t|)} s[i] \neq t[i] \right) + ||s| - |t||}{\max(|s|, |t|)} \dots\dots\dots 2$$

Edit Distance

Edit distance [3, 4] method is utilized for finding the homogeneous attribute between the strings. In Edit distance method apply the operations (insertion, deletion, replacement of a character) to the object from the number of objects to

obtain the other one. Edit distance method is the minimum number of operations applied.

Definition 4 (Edit distance) A set of operation ($Op: S \rightarrow S$), and a cost function $w: Op \rightarrow R$ (R expressing the real numbers). There are number of operations applied on the strings that transform the first string into the second one (and vice versa), edit distance dissimilarity $\delta: S * S \rightarrow [0, 1]$ where is the cost of operations that are less applied for transform s to t .

$$\delta(s, t) = (Op_i)t : Op_n(\dots Op_1(s)) = t \sum_{i=1}^{\min} wOp_i \dots\dots\dots 1$$

String-predicated technique is subsidiary to quantify the similar attribute between the strings if we utilize the very similar string to represent the same concepts [4]. If we utilized the synonyms with dissimilar structure then this technique may be given the low kindred attribute. The result of this technique is more subsidiary if we utilized the similar strings.

Language-predicated technique

Language-predicated technique [4, 7] is withal used to find the similarity between the words. This technique takes names as words e.g. English. Language-predicated technique considered the intrinsic techniques. Language-predicated techniques utilized the various methods which considered strings as sequence of characters. This method takes string as an input and fragmented that string in words and these words may be identified sequence of words. Language-predicated method is predicated upon the Natural Language Processing. Natural Language Processing techniques are acclimated to find the consequential words or terms from the text (Text is composed of strings) and finds the kindred attribute between the ontology entities and their relation.

Constraint-predicated techniques

Constraint-predicated technique [4] is utilized the internal structure of the ontology entities rather than comparing the designations or terms. This technique compares the internal structure of the ontology entities. Latter this is called as relational structure because in this we withal comparing the ontology entities with the other entities which is related [7]. The ontology entities without reference to other ontology entities are the internal structure and the set of relation that an ontology entity has with other is the relational structure. The internal structure is utilized in the database schema matching and relational structure is utilized in matching formal ontology and semantic network.

B. Structure-level technique

Contrary to Element-level techniques, Structure level technique [4, 8] measures the homogeneous attribute of ontology and their instances to compare their relation with other entities or their instances.

Graph-predicated technique

In the Graph-predicated technique measures the homogeneous attribute between the pair of nodes, from the two ontologies predicated on their positions in the graph. Graph-predicated takes the two ontologies as labelled graphs input [2]. In this technique if two nodes are similar then their neighbours is additionally somehow homogeneous. In this technique graph are considered from the ontology and nodes are labelled by the relation name (Mathematically graph of multiple relations of the ontologies). A prevalent homomorphic subgraph of both graphs is finding the homogeneous attribute between the nodes of graph.

Definition 5 (Common homomorphic directed graph) Two directed graph $G = \langle V, E \rangle$ and $G' = \langle V', E' \rangle$ a common homomorphic directed subgraph of G and G' is a graph $\langle V'', E'' \rangle$, such that there exists a pair of homomorphism $f = W \rightarrow V''$ and $g: W' \rightarrow V''$ with $W \subseteq V$ and $W' \subseteq V'$, such that

- $\forall w \in V', \exists u \in V; f(u) = w$ and $\exists v \in V'; g(v) = w$;
- $\forall \langle u, v \rangle \in E \mid w * w, \langle f(u), f(v) \rangle \in E''$;
- $\forall \langle u', v' \rangle \in E' \mid w' * w', \langle g(u'), g(v') \rangle \in E''$.

The homogeneous attribute between the two entities from the ontology can predicate on their relation with other entities in the ontology. There are sundry kind of relations are considered

- r comparing the entities in the direct relation r ;
- r^+ Comparing the entities in the transitive reduction of relation;
- r^+ comparing the entities in the transitive closure of relation r .

Taxonomy-predicated technique

Taxonomy-predicated technique [4] is withal utilized the graph-predicated approach but in this technique only the specialization relation are considered. This type of technique is utilized as a comparison resource for matching classes. The perception after the taxonomy technique is that specialization connect term that are already homogeneous (as a super set or subset of each other), therefore their neighbour are additionally somehow homogeneous.

Instance-predicated technique

Instance-predicated technique [4] is utilized for the comparison of sets of instances of classes. On the substructure of the comparison it is decided that classes are match or not. This technique is relies on the set-theoretic reasoning and statistical techniques. Taxonomy-predicated technique is

withal utilized the graph-predicated approach but in this technique we find the similar attribute between the instances. This technique is very subsidiary for integrating the instances and quantifying the homogeneous attribute between them.

IV. CONCLUSION

At the end of this survey paper, we conclude that matching between the entities plays an important role in ontology matching. In this paper various ontology matching techniques or approaches are discussed like Element-Level technique, Structure-Level technique. In these techniques further discussed the various techniques. In first section discussed the some introduction part, In second section some basic of matching and in third section different kind of ontology matching techniques.

REFERENCES

- [1] Li Huang, Guoxiong Hu, Xinghe Yang, "Review of Ontology Mapping", 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet), Yichang, p.p 537-540, 21-23 April 2012.
- [2] Pavel Shvaiko and Jerome Euzenat, "Ontology matching: state of the art and future challenges" IEEE Transactions on Knowledge and Data Engineering, Vol. No. 25, Issue: 1, p.p 158-176, 2013.
- [3] Bin Ye, Hongyu Chai, Weiping He, Xiaoting Wang, Guangwei Song, "Semantic similarity calculation method in ontology mapping", 2nd International Conference on Cloud Computing and Intelligent Systems (CCIS), Hangzhou, Vol No. 3, p.p 1259-1262, 2012.
- [4] J. Euzenat and P. Shvaiko, "Ontology matching", Springer, 2013.
- [5] A Rodrguez, M Egenhofer, "Determining Semantic Similarity Among Entity Classes from Diferent Ontologies", IEEE Transactions on Knowledge and Data Engineering, Vol. No.15 (2), 442-456, 2003.
- [6] Li Chun-miao, Sun jing-bo, "The Research of Ontology Mnapping Method Based on Computing Similarity ". Science & Technology Information, Vol. No.1, p.p 552-554, 2010.
- [7] Yang Zhe., "Semantic similarity match of ontology concept based on heuristic rules". Computer Applications, Vol. No. 12, Dec. 2007.
- [8] Jiang Huiping, " Information Retrieval and the Semantic Web", IEEE International Conference on Educational and Information Technology (ICEIT), Chongqing, China, Vol No. 3, p.p 461-463, 2010.