Relation based Ontology Matching using Alignment Strategies

[1] Mr. S.Vivekanandan, [2] Mr.P.Marikkannu,

[1]Final M.Tech IT, [2] Head of Department,
Department of Information Technology,
Anna University of Technology, Coimbatore.

ABSTRACT

The set of relation within a knowledge domain will be expressed with a help of Ontology, but data within the knowledge domain get scattered all over its space. To get a most precise result there must be necessary to relate the concepts or keywords within a domain. One of the efficient ways of matching or relating data must be done by implementing ontology within a domain. It aims to make data sharable. Unfortunately, the ontology is widely distributed as well as heterogeneous. The main aim of Ontology matching is to determine the relationship between the concepts and to find the Semantical mappings between two given ontologies. This problem lies at the heart of numerous information processing applications. In order to dilute the problem evolution of Upper ontology have taken place. The upper ontology will act as a base for representing concepts in all the domains, the primary objective is to extract or representing the general concepts throughout the domains based on the ontological structure.

Keywords: - ontology, upper ontology, ontology matching, stemming, RDF, Semantic web, knowledge domain.

1. INTRODUCTION

The World Wide Web is a huge collection of information; the internet is mainly based on interconnection of computers along with collection of interconnected documents and other resources. In day to day life the web search engines play a vital role for to extract information from World Wide Web, hence now there are various kinds of search engines arises, the most advance component nowadays where the search engines are using is nothing but ontology. The main objective of ontology is to relate the objects or concepts this will help the end users to make search more efficient as well as to maximize the result. Ontology typically provides a specific vocabulary that describes a domain of interest and a specification of the meaning of terms used in the vocabulary. The mapping is a primary procedure that followed to relate the concepts, various algorithms and procedures to enhance matching is being proposed, this will help ontology to work fast and also in a better way. The mapping or alignment of ontologies is one of the most important concepts also it acts as a heart of ontology in semantic web search engines. An ontology alignment is the expression of relations between different ontologies. Indeed, alignment results can be an additional support to the visualization of correspondence, the transformation of one ontology into another or the devising of bridge axioms between the ontologies. At present, several systems are available to support users in aligning ontologies, but not lots of relative estimates have been executed and there exists not much support for performing such evaluations. There are various algorithms are available to stem the content inside the database here a new strategy of matching using a stemming algorithm is being proposed.

[1][22] Stemming is the process for reducing inflected (or sometimes derived) words to their stem, base or root form – generally a written word form. The stem need not be identical to the morphological root of the word; it is usually sufficient that related words map to the same stem, even if this stem is not a valid root by itself.
The above mentioned algorithms used to stem the database objects (for example a set of documents that contain stem words). Those words might be common sub-strings, as the "cut" in "cutti" and in "cutting”. To stem in order wise the algorithm try to match it from stems from the data available inside the database, applying various constraints, such as on the relative length of the candidate stem within the word.

2. LITERATURE SURVEY

2.1. ONTOLOGY MATCHING

[1][2][3]In the sequel, an account of the concepts is given which are being used throughout the paper and of the metrics that is used for computing alignments. The terminology proposed in [11] is followed and adopt the same definitions given there, as well as the same symbols within.

The Semantic Web[5][6] is a vision for the future of the Web, in which information is given explicit meaning, making it easier for machines to automatically process and integrate information available on the Web. The Semantic Web will build on XML's ability to define customized tagging schemes and RDF's [10][11][12] flexible approach to representing data. The first level above RDF required for the Semantic Web is an ontology language which can formally describe the meaning of terminology used in Web documents.

If machines are expected to perform useful reasoning tasks on these documents, the language must go beyond the basic semantics of RDF Schema. The OWL Use Cases and Requirement Document provides more detail on ontologies, motivates the need for a Web.

Ontology Language in terms of six use cases, and formulates design goals, requirements and objectives for OWL. OWL has been designed to meet this need for a Web Ontology Language. OWL is part of the growing stack of W3C recommendations related to the Semantic Web.

- XML provides a surface syntax for structured documents, but imposes no semantic constraints on the meaning of these documents.
- XML Schema is a language for restricting the structure of XML documents and also extends XML with data types.
- RDF is a data model for objects ("resources") and relations between them, it provides a simple semantics for this data model, and these data models can be represented in XML syntax.
- RDF Schema is a vocabulary for describing properties and classes of RDF resources, with a semantics for generalization-hierarchies of such properties and classes.
- OWL adds more vocabulary for describing properties and classes: among others, relations between classes (e.g. disjointness), cardinality (e.g. "exactly one"), equality, richer typing of properties and characteristics of properties (e.g. symmetry), and enumerated classes.

The three sublanguages of OWL

OWL provides three increasingly expressive sublanguages designed for use by specific communities of implementers and users.

- OWL Lite supports those users, primarily need a classification hierarchy and simple constraints. For example, while it supports cardinality constraints, it only permits cardinality values of 0 or 1. It should be simpler to provide tool support for OWL Lite than its more expressive relatives, and OWL Lite provides a quick migration path for thesauri and other taxonomies. OWL Lite also has a lower formal complexity than OWL DL; see the section on OWL Lite in the OWL Reference for further details.
- OWL DL supports those users who want the maximum expressiveness while retaining computational completeness (all conclusions are guaranteed to be computable) and decidability (all computations will finish in finite time). OWL DL includes all OWL language constructs, but they can be used only under certain restrictions (for example, while a class may be a subclass of many classes, a class cannot be an instance of another class). OWL DL is so named due to its correspondence with *description logics*, a field of research that has studied the logics that form the formal foundation of OWL.
- OWL Full is meant for users who want maximum expressiveness and the syntactic freedom of RDF with no computational guarantees. For example, in OWL Full a class can be treated simultaneously as a collection of individuals and as an individual in its own right. OWL Full allows an ontology to augment the meaning of the pre-defined (RDF or OWL) vocabulary. It is unlikely that any reasoning software will be able to support complete reasoning for every feature of OWL Full.

2.2. A String Metric for Ontology Alignment

Ontologies are today a key part of every knowledge based system. They provide a source of shared and precisely defined terms, resulting in system interoperability by knowledge sharing and reuse. Unfortunately, the variety of ways that a domain can be conceptualized results in the creation of different ontologies with contradicting or overlapping parts. For this reason ontologies need to be brought into mutual agreement (aligned). One important method for ontology alignment is the comparison of class and property names of ontologies using string distance metrics. Today quite a lot of such metrics exist in literature. But all of them have been initially developed for different applications and fields, resulting in poor performance when applied in this new domain. In the current paper a new string metric for the comparison of names which performs better on the process of ontology alignment as well as too many other field matching problems.

2.2.1. Implementation

A new string metric which is being created by paying special care to each different characteristic of the process in ontology alignment, leads to a metric with very good performance. The new metric is based on the similarity between two entities. The similarity among two entities is related to their commonalities as well as to their differences. Thus, the similarity should be a function of both these features. This feature also appears, sometimes implicitly, in other measures as well.

3. ALGORITHMS FOR ONTOLOGY MATCHING USING ALIGNMENT STRATEGIES

The following algorithms which are implemented compute only the correspondences between the concepts.

1. Stemming Algorithm
2. Three State Matching Algorithm
3. Structural Parallel Match Algorithm
4. Hybrid Match Algorithm

The created reference alignments only match the concepts, and discard the correspondences between individuals and between properties from the alignments computed by the algorithms. The reason of the choice is to find correspondences between properties in a meaningful way; it also required to take their domain and ranging into account.

The algorithms are fully automatic, and hence, they cannot cope with all the possible ontology mismatches. In particular, they do not face conceptualization mismatches, but as per the study there is no automatic tool that copes with them in a satisfactory way. Instead, the algorithms can solve many (even if not all) Term and Definitions and Term mismatches. These mismatches arise when two concepts are synonyms.

The exploitation of WordNet allows the algorithms to create correspondences between c and c^0 if both of them belong to WordNet and they are defined as synonyms there. The usage of upper ontologies may correctly match two concepts that have the same meaning but that were not recognized as synonyms in WordNet.

3.1. STEMMING ALGORITHM

Stemming is the process for reducing inflected (or sometimes derived) words to their stem, base or root form – generally a written word form.

$$(m>0)\ \text{*HIGHNESS} \rightarrow \text{*HIGH}$$

The stem need not be identical to the morphological root of the word; it is usually sufficient that related words map to the same stem, even if this stem is not in itself a valid root.

$$(m>0)\ \text{*HIGH} \rightarrow \text{null},$$

The above relation denotes that the suffix *HIGH should be replaced by the null string if, and only if, the resulting stem has a non-zero measure.

The algorithm is a simple approach towards the conflation which works well in practice and also applicable to a large range of languages. It holds high interest in stemming towards research, rather than deploying as a least component of an information retrieval system.

3.2. THREE STATE MATCHING ALGORITHM

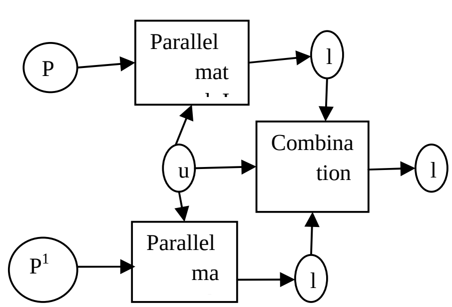

Fig.1. Three State Matching Algorithm

It contains three states as follows:

- Conjunction
- Parallel match
- Combination

Where P, P^1->ontology, l->be the alignment

- **Conjunction**

It produces the alignment obtained by making the union of all the correspondences in a and a , and choosing the correspondence with the highest confidence measure, in case the same correspondence belongs to both a and a .

- **Parallel match**

It computes an alignment between p and p1 by applying substring, n-gram, SMOA, and language-based methods in parallel, and aggregating them. The only external resource used is WordNet, which is given in input to the language based method.

- **Combination**

It computes the alignment a in such a way that a correspondence belongs to a

3.3. STRUCTURAL PARALLEL MATCH ALGORITHM

The structural parallel match uses the three state matching function for computing the alignments, and not the structural parallel match one. The reason for defining a structural parallel match function is to compare homogeneous matching methods.

Thus, in this experiment, the results of parallel match with those of three state match and the results of structural match with those of structural parallel match are compared.

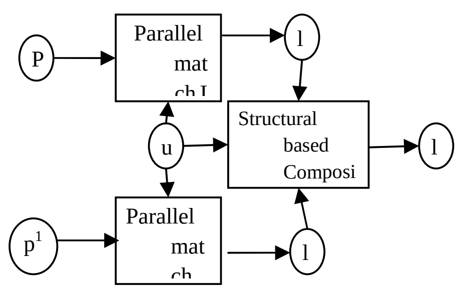

Fig.2. Structural Parallel Match Algorithm.

Fig.2. refers to the structural parallel match process where the heterogeneous ontologies namely p, p^1 are being structurally compared along with the upper ontology u. Where each of the parallel match exhibits alignments l which is integrated by the structural based composition and finally the alignment l is exhibited from it.

3.4. HYBRID MATCH ALGORITHM

A hybrid match algorithm obtained by aggregating the alignment output by the structural match algorithm and the one output by the structural parallel match algorithm.

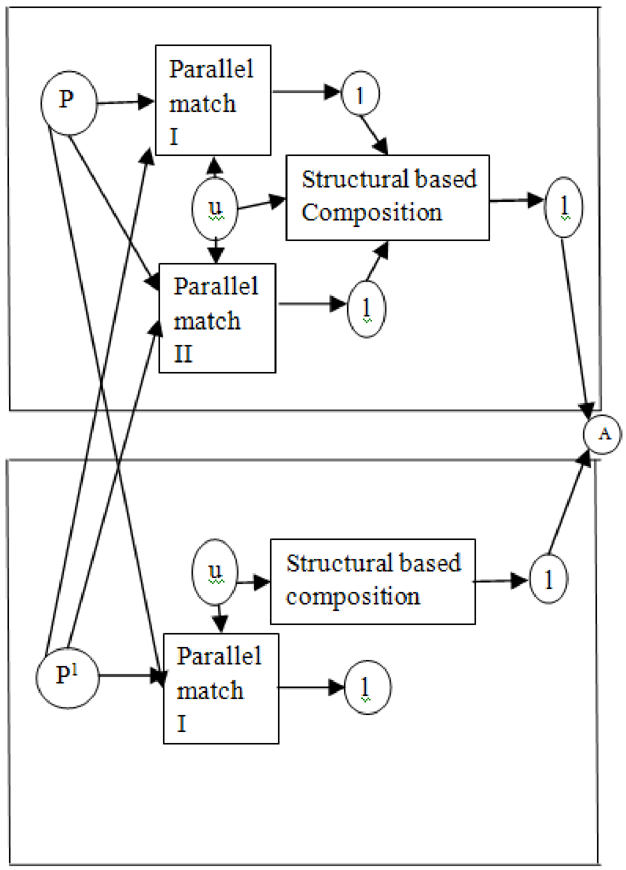

Fig.3. Hybrid Match Algorithm.

Fig.3. refers to the hybrid match process. In this algorithm the previous algorithm's results are being composed that is, the three state match algorithm provides the parallel match combination of the ontologies p, p^1 respectively. Those evaluations are being given as an input to the structural parallel match algorithm where the results of parallel match and the results of structural parallel match are being compared. These two algorithms exhibit alignments which are yet to aggregate in order to yield the maximum efficiency of the process, which is made successful by the hybrid match algorithm.

4. EXPERIMENTAL RESULTS & ANALYSIS

The experimentation is done with the software namely: Windows XP, SQL 2000, Text pad 5, Jdk1.6.0, WordNet and the minimum hardware requirements are Processor: Pentium IV 2.4 GHz ,Hard Disk: 40 GB ,RAM: 256 MB.

Fig.4.Stemmed Words

Fig.4. refers to the stemmed words which are obtained as output from the data sets that are given as input, all the words are stored in a text file for efficient ontology matching. This is obtained by a fully automated alternative to truncation named as stemming Process. This program reduces all words with the

same root to a single form, the stem, by stripping the root of its derivational and inflectional affixes; in most cases, only suffixes that have been added to the right-hand end of the root. The removal of prefixes has been much less studied in the case of English-language retrieval; it is, however, of importance in other languages such as Malay.

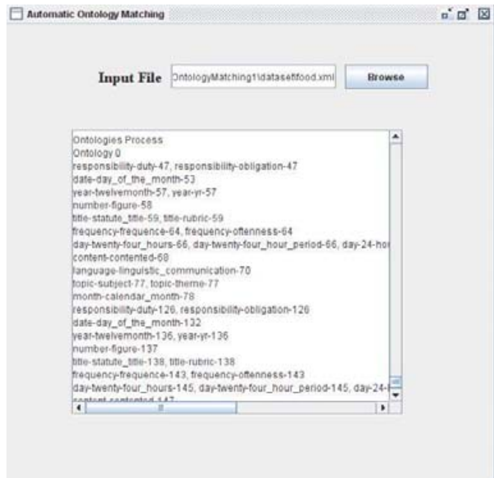

Fig.5.Ontology Process (p, p[1])

Fig.5.refers to the processing of ontologies named p, p[1] respectively. Where the result represents the data set present in the ontology which is being preprocessed through the WordNet and it extracts the Semantical structure of the ontology. The amount of stems tagged by the WordNet domain Domains in ontology are often large in numbers to give a useful hint on the prevalent ontology domains. That is, it gives out the other Semantical meaning of each word and also represent the location of the term present in the ontology.

Onto.	Concepts at the top level(s)
Agent	Activity, Agent, Degree, Gender, Job_title, Language
Bibtex	Entry + (Article, Book, Booklet, Conference, Inbook, Incollection, Inproceedings, Manual, Mastersthesis, Misc, Phdthesis, Proceedings, Techreport, Unpublished)
Bios.	LivingThing, MarineAnimal
Eco.	EcologicalConcepts + (EcologicalEntity, EcologicalEnvironment, EcologicalProcess, EcologicalTerminology, EcologicalTrait)
Food	ConsumableThing + (EdibleThing, PotableLiquid, MealCourse, Meal)
Geo.	Classification, GeographicArea, InstallationType, Location, Status
HL7.	RBAC_Reference_Model_Elements, Human, Industry, Organizational_Resources, Organization, Organizational_Tax_Category
Ka	Object + (Event, Organization, Person, Product, Project, Publication, ResearchTopic)
MPEG7	MPEG7Genre + (Animations-special_effect, Drama, Enrichment, Entertainment, Information, Movies, Music)
Rest.	Atmosphere, Beverage, Catering, Cuisine, DatePeriod, Dish, Facility, Meal, OpeningPeriod, Rating, Restaurant, RestaurantGroupParty, RestaurantSeating, Review, Show, SpecialFeature, TimePeriod
Resume	Accomplishment, Address, Award, Career, ContactInfo, Degree, Education, Experience, ExpertiseArea, Industry, Knowledge, Name, Organization, Patents, Person, Publication, Resume, Title, ValuePartition
Space	Spatial_entity + (Geographical_feature, Geopolitical_entity, Place)
Subj.	AppliedSciences, ArtsAndHumanities, Business, History, InterdisciplinaryStudies, Law, Sciences, SocialSciences
Topbio	Domain_entity + (Refining_entity, Self_standing_entity)
Travel	MetaObject + (Domain, DomainIndependent)
Vac.	CBR_DESCRIPTION, CBR_INDEX, CBRCASE
Vert.	Blood, Body_proper, Brain, Bronchus, Cardiac_valve, Half_heart, Head, Heart, Laterality_selector_value, Left_laterality_value, Limb, Liver, Lobe, Lung, Neck, Pericardium, Right_laterality_value, Stomach, Trachea, Trunk

Table.1. Analysis of the Ontology domain

Table.1. provides an overall description of the ontology domains. For extracting information on the ontology domains precisely in an automatic way, there lies three various approaches: 1) a statistical analysis of stems occurred in the stemmed words; 2) the revealing of WordNet Domains associated with ontology stems; and 3) the analysis of concepts at the upper most level. The frequency of stems in the stemmed words was significant depend on the cases it proclaims.

5. CONCLUSION & FUTURE WORK

The algorithms for ontology matching which may overcome the limitations of other approaches are proposed, the work holds, inclusion of more expressive relations are being made, and the algorithms are made extent in order to properly deal with alignments between attributes and properties, and to perform structural matching methods which advance towards the more significant portion of the framework. Structural alignment will be a greater advantage of all aspects in ontology matching, and it links other than super/sub class frame in the process.

The main future direction is to improve the matching methods along with higher precision. Obtaining a precision higher than the earlier ones is a typical and an efficient result of any automatic ontology matching system.

REFERENCES

[1] Viviana Mascardi, Angela Locoro, and Paolo Rosso, Member, IEEE “Automatic Ontology Matching via Upper Ontologies: A Systematic Evaluation”
[2] Z. Aleksovski, M.C.A. Klein, W. ten Kate, and F. van Harmelen, “Matching Unstructured Vocabularies Using a Background Ontology,” Proc. Int’l Conf. Knowledge Eng. and Knowledge Management (EKAW ’06), pp. 182-197, 2006.
[3] Z. Aleksovski, W. ten Kate, and F. van Harmelen, “Exploiting the Structure of Background Knowledge Used in Ontology Matching,” Proc. Int’l Workshop Ontology Matching (OM-2006),P. Shvaiko, J. Euzenat, N. Noy, H. Stuckenschmidt, R. Benjamins, and M. Uschold, eds., 2006.
[4] Am. Nat’l Standard, “KIF Knowledge Interchange Format—Draft Proposed American National Standard (dpANS),” Technical Report NCITS.T2/98-004, 1998.
[5] L. Bentivogli, P. Forner, B. Magnini, and E. Pianta, “Revising WordNet Domains Hierarchy: Semantics, Coverage, and Balancing,” Proc. Workshop Multilingual Linguistic Resources (MLR ’04), pp. 101-108, 2004.
[6] P.A. Bernstein, A.Y. Halevy, and R. Pottinger, “A Vision of Management of Complex Models,” SIGMOD Record, vol. 29, no. 4, pp. 55-63, 2000.
[7] P. Bouquet, L. Serafini, and S. Zanobini, “Peer-to-Peer Semantic Coordination,” J. Web Semantics, vol. 1, no. 2, pp. 81-97, 2005.
[8] E. Brill, S. Dumais, and M. Banko, “An Analysis of the Askmsr Question-Answering System,” Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP ’02), 2002.
[9] N. Casellas, M. Bla´zquez, A. Kiryakov, P. Casanovas, M. Poblet, and R. Benjamins, “OPJK into PROTON: Legal Domain Ontology Integration into an Upper-Level Ontology,” Proc. On The Move Federated Workshops (OTM ’05), R. Meersman et al., eds., pp. 846-855, 2005.
[10] H.H. Do, S. Melnik, and E. Rahm, “Comparison of Schema Matching Evaluations,” Proc. Web, Web-Services, and Database Systems: NODe 2002 Web and Database-related Workshops,A.B. Chaudhri, M. Jeckle, E. Rahm, and R. Unland, eds., pp. 221-237, 2002.
[11] J. Euzenat and P. Shvaiko, Ontology Matching. Springer, 2007.
[12] A. Gangemi, N. Guarino, C. Masolo, A. Oltramari, and L. Schneider, “Sweetening Ontologies with DOLCE,” Proc. Int’l Conf. Knowledge Eng. and Knowledge Management (EKAW ’02), A. Go´mez- Pe´rez and V.R. Benjamins, eds., pp. 166-181, 2002.
[13] A. Gillett, “Using English for Academic Purposes,” A Guide for Students in Higher Education, School of Combined Studies, Univ. of Hertfordshire, 2009.
[14] F. Giunchiglia, P. Shvaiko, and M. Yatskevich, “Semantic Schema Matching,” Proc. Int’l Conf. Cooperative Information Systems (CoopIS ’05), Distributed Objects Applications (DOA ’05), and Ontologies Databases and Applications of Semantics (ODBASE ’05), R. Meersman, Z. Tari, M.-S. Hacid, J. Mylopoulos, B. Pernici, O ¨. Babaoglu, H.-A. Jacobsen, J.P. Loyall, M. Kifer, and S. Spaccapietra, eds., pp. 347-365, 2005.
[15] F. Giunchiglia, P. Shvaiko, and M. Yatskevich, “Discovering Missing Background Knowledge in Ontology Matching,” Proc. European Conf. Artificial Intelligence (ECAI ’06), G. Brewka, S. Coradeschi, A. Perini, and P. Traverso, eds., pp. 382-386, 2006.
[16] J. Gracia, V. Lopez, M. d’Aquin, M. Sabou, E. Motta, and E. Mena, “Solving Semantic Ambiguity to Improve Semantic Web Based Ontology Matching,” Proc. Int’l Workshop Ontology Matching (OM-2007), P. Shvaiko, J. Euzenat, F. Giunchiglia, and B. He, eds., 2007.
[17] P. Grenon, B. Smith, and L. Goldberg, “Biodynamic Ontology: Applying BFO in the Biomedical Domain,” Ontologies in Medicine, D.M. Pisanelli, ed., pp. 20-38, IOS Press, 2004.
[18] T.R. Gruber, “A Translation Approach to Portable Ontology Specifications,” Knowledge Acquisition, vol. 5, pp. 199-220, 1993.
[19] A. Hameed, A.D. Preece, and D.H. Sleeman, “Ontology Reconciliation,” Handbook on Ontologies, S. Staab and R. Studer, eds., pp. 231- 250, Springer, 2004.
[20] J. Hammer and D. McLeod, “An Approach to Resolving Semantic Heterogeneity in a Federation of Autonomous, Heterogeneous Database Systems,” Int’l J. Intelligent and Cooperative Information Systems, vol. 2, no. 1, pp. 51-83, 1993.
[21] J.A. Hendler, “Agents and the Semantic Web,” IEEE Intelligent Systems, vol. 16, no. 2, pp. 30-37, Mar./Apr. 2001.
[22] M.F. Porter, “An Algorithm for Suffix Stripping,” Program, vol. 14, no. 3, pp. 130-137, 1980.