Ontology Hierarchies Matching by Lattices Alignment*

Hu He[12] and Du Xiaoyong[12]

[1] School of Information, Renmin University of China, 100872, Beijing, China
[2] Key Lab of Data and Knowledge Engineering, MOE, 100872, Beijing, China
{hehu,duyong}@ruc.edu.cn

Abstract. Ontology matching determines the correspondences between concepts and relations of related ontologies. In this paper, we put forward an ontology hierarchies matching approach based on lattices alignment. The proposed lattice-based matching algorithm can be utilized not only in matching processes between two ontologies, but also in annotation processes between an ontology and its corresponding resources. Experiments on spatiotemporal ontology annotation have been carried out which shown the applicability of the approach.

1 Motivations

Ontologies are formal, explicit specifications of share conceptualizations [1]; they provide a formal way to describe concepts and their relations for a specified domain. The Semantic Web vision [2] has greatly promoted people's interest in ontologies. More and more ontologies are put forward by different groups and individuals. As many ontologies for the same domain appearing on the Web, a quantitative evaluation method is needed to discriminate between these ontologies so that we can find the most appropriate one for specified applications.

Large-scale use of ontologies in knowledge discovery and semantic web has stimulated automatic ontology learning and population, with various machine-learning methods applied in these efforts. The evolution of these ontologies also needs to be evaluated quantitatively so that a good evolution can be distinguished from a bad one.

We advocate a lattice representation and assessment algorithm for comparing ontology hierarchies quantitatively in the paper. This approach is based on lattice alignment and can be used for ontology matchmaking, clustering, comparisons and annotations in Semantic Web enabled applications. The proposed lattice based metric can be used both in ontologies matching and ontology annotation between ontologies and their corresponding resources. It has unique advantages comparing with existing measures and algorithms as it provides a unified method which takes into account not only concepts but also the relations between concepts in ontologies.

* This work is supported by China NSFC programs (60496325, 60573092, 60603020).

2 Related Work

Maedche [3] et. al. consider ontologies as two-layered systems, consisting of a lexical and a conceptual layer. They use a set of ontology similarity measures and compare two ontologies separately with concepts and relations. OntoMetric [4] gets for every candidate ontology a quantitative measure of its suitability using: a multilevel framework of 160 characteristics that describe the ontology domain. The specialization of the characteristics and the assessment of the criteria of a particular ontology require considerable effort and thus limit its application for novel users. To use the method, the engineer must compare the importance of the project objectives, and study carefully the characteristics of ontologies. Brewster [5] et. al. use a vector-space model of instances (terms) in a corpus and an ontology to give a measure of the "fit" between the ontology and the corpus (domain of knowledge). The method proposed by Brewster is straightforward and easy to use. But it has some drawbacks too, the method loses structure relations information and uses only the lexical information.

The lattice metric in [7] extends Brewster's works and convert ontologies to be evaluated into lattice structures; based on an algorithm for finding identical concepts, we align two lattices to the same dimension; a traditional Vector Space Model (VSM) [6] can be used afterwards to measure the differences quantitatively. Moreover, we will show that the lattice metric approach can be utilized not only in matching processes between two ontologies, but also in annotation processes between an ontology and its corresponding resources.

3 Lattice based Ontology Hierarchies Matching

3.1 Ontology Lattice

Definition 1 Hierarchy $H(S, \leq)$: Suppose $(S, \prec)$ is a partially ordered set. A hierarchy $H(S, \prec)$ for $(S, \prec)$ is the Hasse diagram for $(S, \prec)$, which is a directed acyclic graph whose set of nodes is S and has a minimal set of edges such that there is a path from u to v in the Hasse diagram iff $u \prec v$.

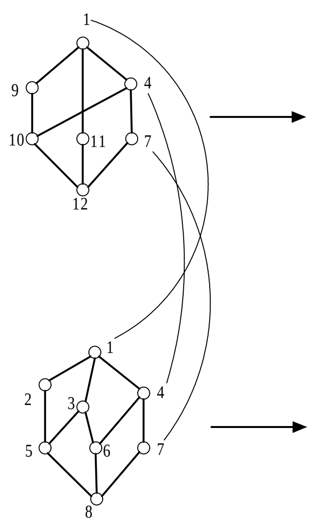
Definition 2 Ontology $O(C, R, H)$: A ontology is represented as $O(C, R, H)$, where C is a set of concepts $\{c_1, c_2, ..., c_i\}$, R is a set of relationships $\{r_1, r_2, ..., r_j\}$, and H is a set of hierarchies H(C,r). There is a root in H(C,r) which is the most abstract concept in C.

Definition 3 Concept Depth $depth(c)$: Define the depth of a concept c node(denoted as depth(c)) in a hierarchy H(C,r) of ontology $O(C, R, H)$ is the number of edges on the path from the root of O to that concept node.

Definition 4 Ontology Lattice: For any particular domain $\mathcal{D}$, and a hierarchy relation $\mathcal{H}$, we use $\prec$ to represent the $\mathcal{H}$ relation: for any two concepts C_1, C_2 satisfies $\mathcal{H}$ (C_1, C_2). We have $C_1 \prec C_2$. Then ($\mathcal{D}$, $\prec$) forms an ontology lattice. Detailed information about lattice construction and alignment algorithm is referred to [7].

3.2 VSM Representation

In most cases, the matrices of two ontologies (or resource and ontology in annotation process) to be matched are not in the same dimensions; to use matrix based comparison methods such as VSM, the matrices must be transformed into the comparison space of the same dimension. After the matrices being transformed into the same dimension space(in Fig. 1), the traditional model of VSM can be used. The measure of two Ontology lattices A and B is formulated as: Diff(A,B)=(Vect(A) · Vect(B)) / |Vect(A) · Vect(B)|. The similarity between lexical entries follows the edit distance formulated by Levenshtein[8] which is a well-established method for weighting the difference between two strings.



	1	2	3	4	5	6	7	8	9	10	11	12
1	0	0	0	1	0	0	0	0	1	0	1	0
2	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0
4	1	0	0	0	0	0	1	0	0	1	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	1	0	0	0	0	0	0	0	1
8	0	0	0	0	0	0	0	0	0	0	0	0
9	1	0	0	0	0	0	0	0	0	1	0	0
10	0	0	0	1	0	0	0	0	1	0	0	1
11	1	0	0	0	0	0	0	0	0	0	0	1
12	0	0	0	0	0	0	1	0	0	1	1	0

	1	2	3	4	5	6	7	8	9	10	11	12
1	0	1	1	1	0	0	0	0	0	0	0	0
2	1	0	0	0	1	0	0	0	0	0	0	0
3	1	0	0	0	1	1	0	0	0	0	0	0
4	1	0	0	0	0	1	1	0	0	0	0	0
5	0	1	1	0	0	0	0	1	0	0	0	0
6	0	0	1	1	0	0	0	1	0	0	0	0
7	0	0	0	1	0	0	0	1	0	0	0	0
8	0	0	0	0	1	1	1	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0

Fig. 1: Aligned Lattices and their VSM Representations.

4 Case Study

4.1 Spatiotemporal Ontology Annotation

Spatial and temporal information constitutes a most elementary part of our everyday life. The representation and reasoning of spatial and temporal knowledge remain an important field in artificial intelligence research. Because much of spatiotemporal information is scattered in free texts, they can not be easily extracted. NLP techniques such as lexical, part of speech, syntactic and semantic reprensentation formalisms are often used in unstructured documents analysis; but the analysis results can be read and understand only by human beings not machines, making it difficult to apply on vast amount of information on the WWW. Spatiotemporal ontology annotation is in urge need to solve these problems.

To prove the correctness and practicability of the lattice-based metric, we choose the spatiotemporal ontology as the experimental data and refine the obtained ontological structure based on Classified Chinese Library Thesaurus, which classifies the controlled vocabulary in particular domains and is adopted widely for organizing literature resources.

The lattice-based matching method is applied to evaluate the spatiotemporal ontology and its corresponding resources; for a particular spatiotemporal resource, we align the resource to the spatiotemporal ontology and construct the lattice structure representation as depicted in Fig. 2; then the resource lattices are compared with the ontology lattice to determine the most appropriate annotation concept, which is the annotation result of the current resource. In practice, we use concept depth value of 3 and choose the maximal matching concept from the ontology as the annotation result for the resource.

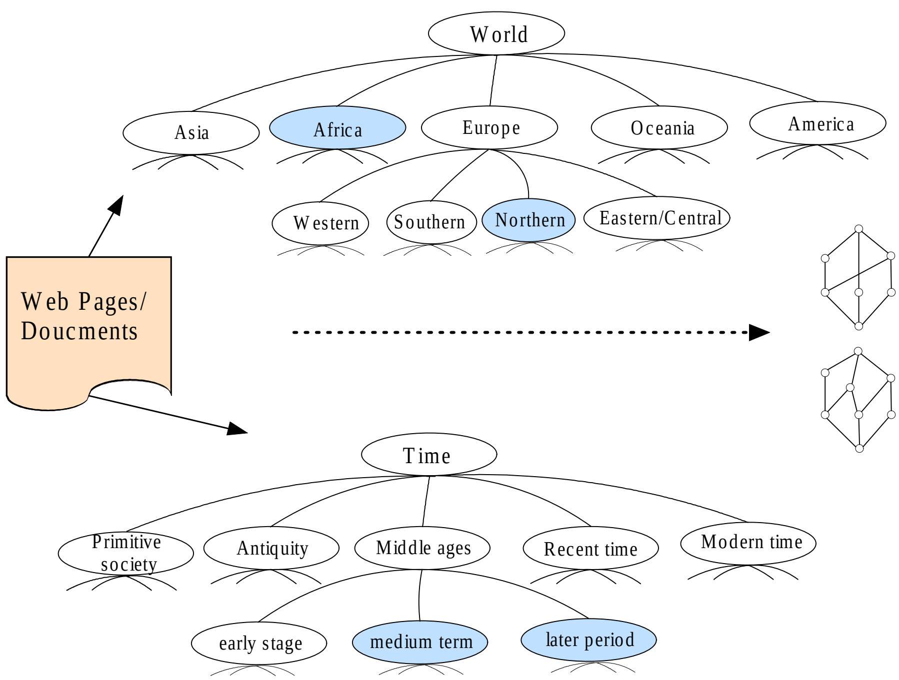


Fig. 2: The Resource Lattice Construction.

4.2 Algorithm Efficiency

Let n be the number of **A**'s nodes, and m be the number of **B**'s nodes; the matrix alignment algorithm need n*m loops of **Find-Identical-Concepts**. Let k be the average number of children nodes for **A** and **B**, the lattice comparison complexity will be $n \times m \times 2k$. We can see the algorithm is polynomial in time complexity, which indicates its' efficiency in real world applications.

5 Conclusion

We propose a unified lattice based approach for ontology hierarchies matching tasks and ontology annotation tasks. A unique quality of this measure method is that it combines the concepts and hierarchy relations into a unified structure: a lattice. By aligning two different lattices, the traditional vector space model can be used in the matching processes. The presented lattice alignment algorithm can be utilized not only in matching processes between two ontologies, but also in annotation processes between an ontology and its corresponding resources. Experiments on spatiotemporal ontology annotation have been carried out which shown the applicability of the approach.

References

1. Gruber T, (1995), Towards principles for the design of ontologies used for knowledge sharing, International Journal of Human-Computer Studies 43 (5/6): 907-928
2. Tim Berners-Lee, James Hendler and Ora Lasilla, (2001), The Semantic Web, The Scientific American.
3. A. Maedche and S. Staab, Measuring Similarity between Ontologies,(2002), In: Proc. Of the European Conference on Knowledge Acquisition and Management (EKAW 2002), Madrid, Spain
4. Lozano-Tello, A. and Gmez-Prez, A., (2004), ONTOMETRIC: A Method to Choose the Appropriate Ontology, Journal of Database Management, Special Issue on Ontological analysis, Evaluation, and Engineering of Business Systems Analysis Methods, Volumen: 15(2)
5. Christopher Brewster, Hartith Alani, Srinandan Dasmahapatra, and Yorick Wilks, (2004), Data driven ontology evaluation, in Proceedings of LREC 2004, Lisbon
6. G. Salton, M. Lesk, (1968), Computer evaluation of indexing and text prcessing, Journal of the ACM, 12(1), pp.8-36
7. H. Hu, X. Du, (2007), A Lattice Metric for Evaluating Ontology Hierarchies, In: Proceedings of International Conference of Machine Learning and Cybernetics (ICMLC 2007), pp.3880-3883, Hongkong
8. Levenshtein, I.V., (1966), Binary codes capable of correcting deletions, insertions, and reversals. Cybernetics and Control Theory, 10(8):707-710