International Journal of Database Theory and Application
Vol.9, No.7 (2016), pp.257-268
http://dx.doi.org/10.14257/ijdta.2016.9.7.22

Conceptual Cluster-based Large-scale Ontology Compression Approach

Yang Feng [1] and Qin Ziting [2]

1(College of Informational Engineering, Jilin Engineering Normal University, Changchun China
yangfeng06@mails.jlu.edu.cn)
2(College of Computer Science and Technology, Jilin University, Changchun, China)
18844548892@163.com

Abstract

With the development of semantic web and ontology application, there is a large number of ontology whose scale is large and the structure is complex in different fields. The existing mapping method and mapping system perform well when dealing with the mapping between a lightweight small ontology. However, when comes to the large-scale ontology, it is full of challenges to the methods and systems。 To this end, this paper proposes a method of ontology compression based on conceptual cluster to compress. Firstly, it calculates the semantic similarity and semantic correlation of ontology concepts with the DICE coefficient method and the information entropy technology to get semantic relation. Secondly according to the semantic relations, it carries on the conceptual cluster in the concept space so that the concept of semantic relations closely together in groups. The concept of cluster in space is reduced, and the "noise concept" which is independent of the mapping is removed, and the purpose of the large-scale ontology compression is realized. Experimental results show that the method is so effective that it can compress the volume of large-scale ontology in the mapping problems.

***Keywords**: Ontology; Ontology mapping; Conceptual Cluster; Compression*

1. Introduction

Ontology is increasingly applied in solving the inconsistency problems of data and information in semantic WEB. With the increasing demand of semantic interoperability, in order to meet the need of more semantic applications, ontology which is larger and more complex in structure appears such as SNOMED-CT [1], GALEN [2], FMA [3] and NCI-Thesaurus [4] in the biomedical domain. Compared with the lightweight ontology, the number of entities in the large-scale ontology is more, the hierarchical structure is deeper, and the constraint relationship between entities and entities is more complex.

As for a given ontology mapping method or system, the input is just the heterogeneous ontology in the same field. When the number of the concepts is very large, there will be a sharp drop in performance just as Qian Zhong [5] pointed out. In addition, the large-scale ontology contains numerous concepts that have nothing to do with matching tasks, such as noise which interfere a lot with discovering semantic relations. In consequence, some characteristics don`t contribute to discovering for mapping.

The traditional ontology mapping method performs well only when scale is lightweight and small [6]. In a word, it has already become a hot topic in the field of discovery for large-scale ontology mappings. In order to optimize the large-scale ontology mapping, this paper presents a compression method which reserves the useful elements related to mapping discovery and get rid of the meaningless noise to by exploring the semantic

ISSN: 2005-4270 IJDTA

relations among ontology concepts. Reduce the scale and the complexity of mapping problems.

2. Ontology and Ontology Mapping

2.1. Ontology

Ontology concept originates from philosophy, ontology is "a systematic description of objective, concerned with abstract entities of objective reality." In the domains of Artificial Intelligence, Knowledge System and Information System, ontology is "a formal, clear and detailed explanation of sharing concept system" [7]. Ontology provides a sharing word list, that is object type or concept and its attributes and relationships exist in particular domains [8]. [9] uses taxonomy organize ontology, and summarize 5 basic Modeling Primitives, proposes definition of ontology.

Definition 1. Ontology

Ontology O= { C, R, F, A, I }, where C is class or set of concepts, c is concept ($c\in C$), which indicates everything, like work specification, function, behavior, strategy and reasoning process; R is relationship, the interaction between concepts in domain, defines as subset of n-dimensional Cartesian product formally: $R:C_1\times C_2\times\ldots\times C_n$, r is relationship ($r\in R$), basic relationships contain: subclass-of, part-of, kind-of and attribute-of; F is function, a kind of special relationship. Formal definition is $F:C_1\times C_2\times\ldots\times C_{n-1}\rightarrow C_n$, such as Mother-of is a function, Mother-of(x, y) means y is mother of x; A is axiom, represents tautological assertion, like concept B belong to the range of concept A; I is set of instances, i is instance ($i\in I$).

2.2. Ontology Heterogeneity

The development of Semantic Web makes more and more ontologies. Different organizations define ontology to serve their own applications. Because there is no unified standard for ontology construction, the content and structure of these ontologies have differences. It becomes a balk to knowledge sharing and multiplexing. Ontology heterogeneity exists in Semantic Web widely, and registers as various ways. For example: different concepts come from two ontologies and contained the same semanteme are defined as different name labels, this status is called name heterogeneity; for another example: If their semantemes are very different, concepts are defined as a same name labels, here becomes a semantic heterogeneity. In addition, same parent class has different numbers of sub-classes in different ontologies, namely structure heterogeneity. To achieve distributed information integration and knowledge multiplexing, it must find semantic relationships between different ontologies by ontology mapping, and then to solve Ontology heterogeneity.

2.3. Ontology Mapping

Ontology mapping is a procedure, it uses two independent ontologies as input, and creates interrelated semantic relationship of all the elements (concept, relation, entity) in the two ontologies [10].

Definition 2. Ontology Mapping

Ontology Mapping is to find semantic relationship between two ontologies, mapping function [16] expressed as $Map(\{e_{1i},e_{2j}\},O_1,O_2)=f$ .Two given ontologies O_1 and O_2, mapping from ontology O_1 to O_2 means every entity in ontology O_1 finds a appropriate entity in ontology O_2, and give their corresponding relation. This corresponding relation is usually determined by similarity between the two entities. O_1, O_2 are called source ontology and target ontology respectively. Here $e_{1i} \in O_1$, $e_{2j} \in O_1$, and $\{e_{1i}\} \xrightarrow{Map} \{e_{2j}\}$. They are all element sets. f is a kind of mapping type or null. When f is null, means there is no mapping relationships between { e_{1i} } and { e_{2j} }.

3. Related Work

In order to deal with the large-scale ontology mapping, researchers put forward some methods at present among which the following three methods are the most representative. Lopsided ontology mapping refers to the wide disparity in scale - one is large while the other is quite small. Gass function [11-12] method extracts the sub ontology with the small scale in the large-scale ontology, and tries to find the mapping relationship between the sub ontology and the small ontology which result in the final mapping. This is a method to deal with the mapping relationship between the lopsided ontology whose input is the ontology that different in scale. Frédéric Fürst and Francky Trichet presented a method to deal with large scale ontology mapping by axiom [13-15]. With the help of Ontology Conceptual Graphs Language and the corresponding processing tools: TooCom, the two semantic relations of the axioms are expressed as graph structure. And then this method expands the reasoning on this graph according to the theory and technology of graph theoretic to find the analogies. This method is very sensitive to ontology structure, as well as other graph-based methods. The large scale ontology mapping method based on data field is proposed by QianZhong from Tsinghua University [16]. According to the mapping relationship between the sub-ontology from related fields, the whole mapping discovery is completed thanks to that the scale of problem is reduced and the task is more efficient. However, the users are required to have a certain understanding of data and set some certain parameters and thresholds in the mapping process.

4. Large-Scale Ontology Compression Based on Conceptual Cluster

4.1. Method overview

According to the definition and description of ontology, the ontology gathers the terms and vocabulary from a certain field to describe the semantic information by the relationship between them. These terms or words are the concepts, which is a unified abstract description of a set of examples as the core of ontology. In the whole concepts, there are not only the useful concepts for the final mapping, but also the ones without any semantic relations. Those irrelevant concepts are to be blame for a complex and low-speed result. For example, there is ontology waiting for mapping, O_1 and O_2, which contain m and n concepts respectively. That mains a m*n-time calculation of similarity is needed result in m*n dimensions of the similarity matrix. Although there are no significant effect on the quality and efficiency of the mapping for small-scale and lightweight ontology, it comes to be a considerable problem for large-scale ontology mapping tasks. It is high time to reduce the scale by filtering all the concepts logically and involving the meaningful concepts in the mapping method.

4.2. Ontology Concept Semantic Relation Measurement

Ontology concept semantic relationship is made up of two aspects - the semantic similarity and the semantic correlation between concepts which together constitute all the semantic relations. All of them must be examined when deal with the ontology mapping. Only in this way can we measure the concept relations completely and provide the accurate information for the mapping. In particular, in the ontology mapping, element similarity computation is the most important step in the mapping. Element similarity computation is used to restore to the vocabulary prototype after dealing with the text content of the ontology element based on which the similarity is measured to get a value range of [0-1] . As two extreme cases, the similarity is 0 when two elements cannot be replaced in the upper and lower structures of different ontology, in contrast, the similarity is 1.

Matching technology is the main method of ontology element similarity measurement including BM [17], Sunday [18], SundayNew [19], ABNDM [20] *etc.* which is selected for different applications. According to the characteristics of the research, we use the DICE coefficient to calculate the semantic similarity of ontology concepts. Dice coefficient which is put forward by Lee Raymond Dice in 1945 in the literature [21] is used to calculate the similarity between two different words in the field of information retrieval. It describes the similarity between the words by the number of matched lexical and the amount of all the lexical which is defined by [0-1]. This approach is extraordinary fast especially when dealing with large scale data. Dice coefficient is used in text mining [22], information integration [23] and biological information [24]. More details of Dice coefficients is given by literature [20] according to which we define the concept of ontology as follows:

Definition 3: Ontology Concept Similarity Sim_{dice}

Two ontology O_1 , O_2 and the similarity of concepts $c_{1i}(c_{1i} \in O_1)$, $c_{2j}(c_{2j} \in O_2)$ are given. And the similarity is double ratio of the same number of characters and the total number of two characters of t the number of. That is:

$$Sim_{dice}(c_{1i}, c_{2j}) = \begin{cases} \dfrac{2 \times Comm(c_{1i}, c_{2j})}{L(c_{1i}) + L(c_{2j})} \\ 1, \quad if\ \ c_{1i} = c_{2j} \text{ or } c_{1i}, c_{2j} \in \text{equivalence class of vocabulary} \end{cases}$$

Among them, $Comm(c_{1i}, c_{2j})$ is the number of the same characters in c_{1i} , c_{2j} . $L(c_{1i})$, $L(c_{2j})$ is the character length of c_{1i} , c_{2j}

The semantic relationship cannot be fully expressed only with the ontology similarity. For example, there are two concepts "vehicle" and "car". The similarity between them is very low as the similarity of Dice coefficient is only 0.2.But these two concepts have a very high degree of similarity. In order to fully express the semantic relations, it should take into account the similarity between the ontology concepts in addition to the similarity of concept labels. Concept correlation is another important index to measure the semantic relations which reflects the correlation between the two concepts. It can be expressed by a real number between [0-1]. Note that the relationship between concepts in the ontology is reflected by their respective properties. There are two types of ontology attributes: object attributes and data attributes [25]. The object property connects two different concepts, while the data attribute associates a specific data type or character type. As a result, the concept attributes become an important aspect to measure the ontology concept correlation. In this chapter, we use the information entropy to measure the correlation degree between concepts.

Clausius first proposed the concept of "entropy (Entropy)", in order to express the second law of thermodynamics in the form of the increase of the system entropy[26]. "Entropy" expresses the disorder degree of a system which applied in many fields, such as: Life Science, information theory, physics, control theory, *etc.*

[27-30].In information theory, it is called "information entropy" which is a measure and a symbol of the amount of information as a description of uncertainty of the system [31]. As mentioned before, in essence, the complex relationship between ontology concepts is expressed by the specific attributes they have which associates two different entities in the ontology. This natural characteristic makes it an important way to measure the relative degree of ontology concept. In this chapter, the information entropy of attribute is measured to get the semantic correlation. The relevant definitions are as follows:

Definition 4: concept of the property collection is expressed as:

$$S_p(c) = \{p_1 | 1 \le i \le n\}$$

c represents a concept in the ontology, p_1 is a property of the concept.

Definition5: Total Rate of the Property

For a given concept $c_{1i}(c_{1i} \in O_1), c_{2j}(c_{2j} \in O_2)$, property of p_1 is the ratio of the times it appears and the total property of the two concepts, *i.e.*:

$$f(p_i) = \begin{cases} \dfrac{k}{n+m} & (p_i \in S_p^1) \wedge (p_i \in S_p^2) \\ & 0 \end{cases}$$

The S_p^1 and S_p^2 represent the attributes of c_{1i} and c_{2j} respectively, which contain n and m elements; k represents the times that p_1 appears in S_p^1 and S_p^2.

Definition 6: Attribute Information Entropy

The attribute information entropy is defined as:

$$H(p_i) = -\sum_{p_i \in s_p} f(p_i) \times \log_2 f(p_i)$$

p_i represents an attribute, the S_p is a collection of attributes. When the $H(p_i)$ is bigger, the attribute information entropy is bigger and useful information the attribute takes along is more.

Definition 7: Semantic Correlation for Information Entropy of C_1 and C_2:

$$Sim_H(c_{1i}, c_{2j}) = \sum_{i=1}^{n} H(p_i)$$

In the above, the semantic relationship between c_{1i} and c_{2j} can be obtained.

Definition 8: Semantic Relationship between C_1 and C_2 is:

$$Sim(c_{1i}, c_{2j}) = \alpha Sim_{dice}(c_{1i}, c_{2j}) + \beta Sim_H(c_{1i}, c_{2j})$$

α and β are two weight coefficients, which represent the weights of the concept similarity and weight of the similarity of information entropy and $\alpha + \beta = 1$. In practice, the value of α and β is adjusted according to the structural characteristics.

4.3. Compression based on Concept Cluster

We get the semantic relations in the last section based on which, we will focus on the compression of ontology.

Definition 9: Conceptual Cluster

S is a set constituted by all the concepts in O ; D_1 and D_2 are the subsets of S_1 and S_2 Respectively and $S_1 \in O_1$, $S_2 \in O_2$. Then $D = D_1 \cup D_2$ is called conceptual cluster.

Definition 10: Concept Aggregation

A process for subset D_1 and D_2 from S_1 and S_2 to product the conceptual cluster D .

Definition 11: Similarity between Cluster

For two different clusters D_p and D_q including m and n concepts, the similarity is the average value. That is:

$$Sim(D_p, D_Q) = \sum Sim(c_m, c_n) \big/ (m+n)$$

c_m and c_n are concepts and $c_m \in D_1$, $c_n \in D_2$.

Definition12: Similarity in Cluster

For a conceptual cluster D, the similarity is defined as:

$$Sim(D) = \sum Sim(c_{1i}, c_{2j}) \big/ (i+j)$$

Definition13: Conceptual Cluster Coupling

For the two conceptual cluster D_p and D_q, the coupling in the process of concept aggregation is:

$$Coupled(D_p, D_q) = \sum Sim(D_p, D_q) Sim(D_q) Sim(D_p)$$

Here is the algorithm:

Input: concept $c_{1i} (c_{1i} \in O_1)$ and $c_{2j} (c_{2j} \in O_2)$,threshold δ

Output: cluster D which had the most concepts

```
Begin
   If NLSim(c1i,c2j)is the maximum, thenc1i,c2j become a cluster D

   If Coupled(Dp ,Dq)< δ
For each Dp
     For each Dq
    Calculate Coupled(Dp ,Dq)
        If Coupled(Dp ,Dq)is the maximum, then Dp ,Dq become a cluster D while the number of cluster is -1
        EndFor
    EndFor
End
```

Select the conceptual cluster D with the largest number of concepts

4.4. Compression Process

The large-scale ontology compression based on conceptual cluster in this chapter is shown in Figure 1 including information processing, semantic relations, conceptual cluster iteration, clustering and candidate mapping. The five steps are as follows:

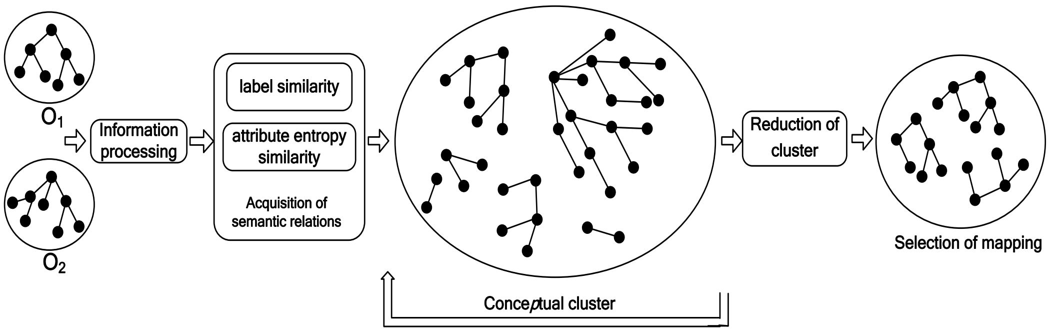

Figure 1. Ontology Compression Architecture

- **Information processing**

Information processing is a necessary step of ontology mapping which is to transform the original representation of the ontology into the form of a convenient semantic relational operation, which may involve complex conversion. Namely: extract the ontology concept labels and concept attributes represented by RDFS or OWL .Then transform into the form which is convenient for finding the semantic relations. Natural language processing (NLP) is often used in the process. In particular, according to segmenting, uniting the size of the written form, deducting abbreviations, removing discarded words, processing plural reduction, the extracted characteristics is changed as a final outcome for the semantic relations.

- **Acquisition of Semantic Relations**

The acquisition of semantic relations is the basis of ontology compression, which is to obtain the semantic relations between ontology concepts, including two aspects: semantic similarity and semantic correlation. In this step, these two aspects are realized by using the DICE coefficient method and the information entropy method respectively. Finally, gather the label similarity and the attribute entropy similarity of the concept as the semantic relations between the two concepts.

- **Conceptual Cluster**

The conceptual cluster is the core part of the method of this chapter whose operation is carried out in the concept space composed of all the concepts dealing with all the objects by compression algorithm. According to the value of the semantic relations, the concept whose semantic relations are closely related gathered as a cluster. That is to say, for the ontology from the same cluster, the semantic relationship is closely, in contrast, it is quite different from the other clusters. This process is not done overnight, it needs to be repeated several times, to get the best results of the cluster.

- **Reduction of Cluster**

After the clustering iteration, it forms a number of conceptual cluster s in the concept space which include a varied number of clusters. The number of concepts determines the size of the cluster. The ontology concepts exist in a large cluster which is meaningful for mapping. At the same time, 'noise concepts' , which is not related to other concepts and independent of the mapping ,are in the small clusters or scattered in the concept space. Cluster reduction is about reducing the size of these small clusters and the scattered concept, in order to achieve the purpose of compress the large-scale ontology of the mapping to reducing the size of the problem solving and improve the efficiency of mapping.

- **Selection of Mapping**

The conceptual cluster after reducing can be used as the input of other mapping methods, and it can be found on the map which depends on the scale of the concept scale. In general, if the heterogeneous ontology is quite large and also close, the volume of the final cluster is large. In turn, if the correlation between the ontology is lower, then the concept of the cluster will be smaller and the final mapping relationship can be accomplished by artificial.

5. Experiments and Evaluation

5.1. Experimental Dataset

In order to verify the validity of the compression method in this chapter, we use the two sets of data in OAEI2014 [32] AGROVOC and NAL. They are a mapping task from OAEI2014: the data set for 'food'. AGROVOC is a dictionary accessed to different languages constructed by Food and Agriculture Organization of the United Nations which contains about 28439 concepts and 3 different types of relationships whose maximum depth of inheritance relations is 8. In [33], we provide different formats of the data set such as: MySQL, TagText and Access Database Microsoft, *etc*. In the experiment we used the OWL language to describe the version. The fragment of the data set is shown in Figure 1, inside which is the concept. The arrows indicate the subclass inheritance relation and subclasses point to the parent class.

NAL (Agricultural Library Agriculture) is a library to construct information service of agriculture by United States Department of Agriculture. It contains more than 42327 concepts and a variety of concepts relations whose maximum depth is 4 [34], providing two languages of NAL: Spanish and English. In the experiment, we chose the English version. Figure 2 shows the segment of the data set.

5.2. Experiment Evaluation

There are two methods to measure the ontology mapping in general: one is the evaluation method using the recall and precision; another is evaluation method based on 'golden standard' set by experts. The method based on the recall and precision is found by the amount of mapping result and proportional relationship between the fact, the method based on 'golden standard' evaluation requires domain experts to evaluate the results of mapping system through the manual construction evaluation criteria. In these two methods, the former is more suitable for the case of a clear mapping. The method proposed in this chapter is intended to reduce the size of the ontology and the result is a collection of ontology concepts . In view of this, traditional assessment methods of recall and precision will not be applied to evaluate the experimental .We should evaluate the experimental results by resetting these two criteria.

First, assuming that we have gotten m clusters expressed as $\{CC_a^u \mid 1 \le u \le m\}$ and experts gives n clusters expressed as $\{CC_p^v \mid 1 \le v \le n\}$ the number of common concepts is $\left|CC_a^u \cap CC_p^v\right|$ for $\{CC_a^u\}$ and $\{CC_a^v\}$:

Therefore the similar concept cluster can be defined as:

$$S(\{CC_a^u\}) = \{CC_p^v\} \qquad if \quad |CC_a^u \cap CC_p^v| \quad \text{is the max}$$

According to the similar concept cluster, the formula of a certain cluster is:

$$Recall' = \frac{\left|CC_a^u \cap S(\{CC_a^u\})\right|}{S(\{CC_a^u\}}$$

The recall is as follows:

$$Precision' = \frac{\left|CC_a^u \cap S(\{CC_a^u\})\right|}{\left|CC_a^u\right|}$$

In the last, we can get the formulas :

$$Recall = \frac{1}{m}\sum_{u=1}^{m}\frac{\left|CC_a^u \cap S(\{CC_a^u\})\right|}{S(\{CC_a^u\}}$$

$$Precision = \frac{1}{m}\sum_{u=1}^{m}\frac{\left|CC_a^u \cap S(\{CC_a^u\})\right|}{\left|CC_a^u\right|}$$

To evaluate the result synthetically, F-measure is:

$$F - measure = \frac{2}{1/Recall + 1/Precision}$$

5.3. Design of Experiment Method

In the experiment, we designed two methods. The first method is to find the mapping relationship between ontology by calculating Semantic similarity and semantic correlation with the gotten semantic relations as a single mapping method to compare with the compressed method.

The second method is to use the ontology compression of this chapter. It carries on the process of the conceptual cluster with the first method of measuring the semantic relations and compresses the scale of the mapping ontology.

5.4. Experimental Results and Analysis

Experimental results are shown in Table1-4.

Table1 gives the first experimental method, that is, only the results of the mapping of semantic relations are used.

Table 1. Experimental Results on Semantic Relations-Only Used

	Precision	Recall	F-measure
AGROVOC to NAL	0.67	0.59	0.62

As can be seen from the Table1, only using a single semantic relationship for mapping is not satisfactory.

The recall rate is only 0.59 and the precision is only 0.67. It shows that, for the large-scale ontology that contains a large number of concepts, properties, instances, constraint relations and classification structures, it is not effective to use a single semantic relationship to map the results.

Table 2. Comparison of Concepts Number

Ontology	Number of Source Ontology	Ontology number after Compression	Ratio
AGROVOC	28439	20417	0.72
NAL	42327	33429	0.79

AGROVOC and NAL are given in the reference map of [35] which contains 9 types of mapping relations, each involving different fields, which contains a varied number of mapping results. After the compression, 11 conceptual clusters are obtained, which contains a varied number of concepts. Table2 and 3 give the concept of their numbers.

Table 3. Concept Number of Clusters

CC_a^1	CC_a^2	CC_a^3	CC_a^4	CC_a^5	CC_a^6
11992	4933	19646	850	850	2041

Continuous Table 3 Concept Number of Clusters

CC_a^7	CC_a^8	CC_a^9	CC_a^{10}	CC_a^{11}	Total
3232	5273	4933	57	39	53846

For the conceptual cluster in the experiment, the number is roughly equivalent to mapping class of the experimental data ,only 2 more than the class, with a less number of concepts of 57 and 39 respectively. Table 4 gives the result of the recall and precision.

Table 4. Experimental Results of Compression

	Precision	Recall	F-measure
AGROVOC to NAL	0.89	0.73	0.80

From Table 4, we can see that the Large Scale Ontology compression is effective with the recall of 0.89, the precision of 0.73 and the F-measure value of 0.80. This shows that it is effective to compress the large scale ontology by the semantic relations to carry out the concept of cluster.

Figure2 gives a comparison of experiments using a single method and a compression method. The recall, precision and F-measure is presented from left to right in the column.

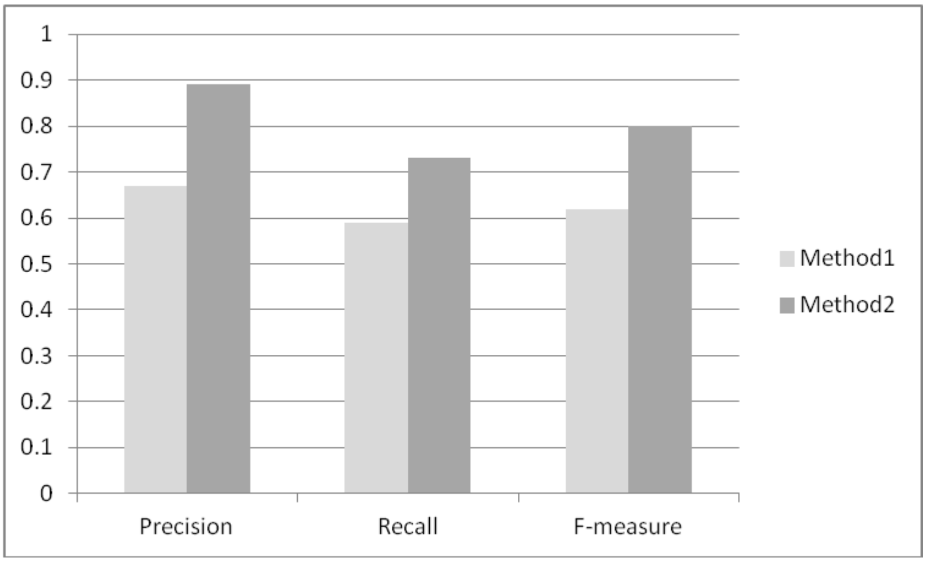

Figure 2. Comparison of Experimental Results

Suggesting that compressing the large-scale ontology by semantic relations in conceptual cluster to process conceptual cluster , can reduce the concepts that has nothing to do with mapping, retain the most concepts related to mapping safely and eliminate influence of ”noise data” so as to improve the quality and effectiveness of the mapping. Experiments show that the method is effective .The concepts of AGROVOC ontology work up to 20417 with a reduction of 8022 compared with 28439 whose compression ratio is 1:0.72. The concepts of NAL ontology work up to 233429 with a reduction of 8898 compared with 42327 whose compression ratio is 1:0.78. A total reduction adding up to 16920 whose average compression ratio is 1:0.75 .And the method of recall and precision reached 0.89 and 0.73 respectively. At the same time, in contrast with the method using only the semantic relations, recall and precision are improved by 0.27 and 0.14.

The experiment does not make a large reduction of the concepts from the ontology. The reason is that the ontology is associated with a high degree of semantic similarity and semantic correlation. If the ontology waiting to map come from two quite different fields, it will be more obvious when the correlation degree is lower. In addition, if the method is carried out on two ontology of different scale, it will also show more obvious effects. At the same time, in order to obtain good precision while taking into account the recall, there are 11 concept cluster in the experiment which is two more than the mapping class in the outcome. However, the number of concepts contained in these two concepts cluster is quite small as X and X, which can be reduced by adjusting the threshold.

6. Conclusion and Future Works

In order to optimize the large-scale ontology mapping, this paper presents a compression method which reserves the useful elements related to mapping discovery and get rid of the meaningless noise to by exploring the semantic relations among ontology concepts. Reduce the scale and the complexity of mapping problems. Experiments show that the method is not only very effective but also apply to process large scale ontology mapping tasks. Aside from striving to improve the accuracy of our approach, in the future, we intend to adjust the method to process m: n mappings.

Acknowledgements

The work of this paper is supported by Natural Science Foundation of Jilin Province in 2016, “Research on large scale ontology mapping mechanism in cloud computing environment”.

References

[1] M. Q. Stearns, C. Price and K. A. Spackman, “SNOMED clinical terms: overview of the development process and project status”, Proceeding AMIA Symp., vol. 662, no. 6, **(2001)**, pp. 503-512.
[2] A. Rector and J. Rogers, “Ontological and practical issues in using a description logic to represent medical concept systems: Experience from GALEN, Reasoning Web, Second International Summer School”, Tutorial Lectures, vol. 4126, pp. 197-231.
[3] C. Rosse and J. L. V. Mejino, “A reference ontology for biomedical informatics: the Foundational Model of Anatomy”, Journal of Biomedical Informatics, vol. 36, no. 6, **(2003)**, pp. 478-500.
[4] J. Golbeck, G. Fragoso and F. Hartel, “The National Cancer Institute’s Thésaurus and Ontology, Web Semantics: Science, Services and Agents on the World Wide Web., vol. 1, no. 1, **(2003)**, pp. 75-80.
[5] Q. Zhong, H. Li, J. Li, G. Xie, J. Tang, L. Zhou and Y. Pan, “A Gauss Function Based Approach for Unbalanced Ontology MatchingSIGMOD’09”, Providence, Rhode Island, USA, **(2009)**.
[6] H. H. Do and E. Rahm, “COMA-a system for flexible combination of schema matching approaches”, in Proceedings of VLDB, **(2001)**, pp. 610-621.
[7] T. R. Gruber, “A Translation Approach to Portable Ontology Specifications”, Knowledge Acquisition, vol. 5, **(1993)**, pp. 199-220.
[8] R. Studer, V. R. Benjamins and D. Fensel, “Knowledge Engineering: Principles and Methods”, Data and Knowledge Engineering, vol. 25, no. 1-2, **(1998)**, pp. 161-197.
[9] A. G. Perez and V. R. Benjamins, “Overview of Knowledge Sharing and Reuse Components: Ontologies and Problem Solving Methods”, In Proceedings of the IJCAI299 workshop on Ontologies and Problem-Solving Methods (KRR5), **(1999)**, pp. 1-15.

[10] E. Rahm and P. A. Bernstein, "A survey of approaches to automatic schema matching", The VLDB Journal, vol. 10, no. 4, **(2001)**, pp. 334-350.
[11] Q. Zhong, H. Li, J. Li, G. Xie and J. Tang, "A Gauss Function based approach for unbalanced ontology matching", In Proceeding of the 2009 ACM SIGMOD international conference on Management of data (SIGMOD'2009), **(2009)**.
[12] F. Shi, J. Li and J. Tang, "Actively learning ontology matching via user interaction", In Proceeding of the 8th International Conference of Semantic Web (ISWC'2009), **(2009)**.
[13] F. Fürst and F. Trichet, "Axiom-based ontology matching: a method and an experiment".
[14] F. Fürst and F. Trichet, "Axiom-based ontology matching", In Proceeding of the 3rd international conference on Knowledge capture (K-CAP'05), pages 195-196, New York, NY, USA, ACM Press, **(2005)**.
[15] F. Fürst and F. Trichet, "Ontology Matching Based on the Comparison of Axioms Represented with Conceptual Graphs", International Conference on Complex, Intelligent and Software Intensive Systems, **(2008)**, pp. 853-858.
[16] Z. Qian, L. Juanzi, T. Jie and Z. Lizhu, "Data Field Based Large Scale Ontology Mapping", Chinese Journal of Computers, vol. 33, no. 6, **(2010)**, pp. 955-965.
[17] Y. Yuefu and S. Liquan, "The Study of Multi-agent Cooperative Framework Based on Pervasive Computing", Journal of Harbin University of Science and Technology, vol. 15, no. 1, **(2010)**, pp. 19-23.
[18] D. M. Sunday, "A Very Fast Substring Search Algorithm", Communications of the ACM, vol. 33, no. 8, **(1990)**, pp. 1,32-14
[19] L. Yupeng, M. Chunguang, L. Shui, L. Lemao and Z. Shilei, "The Summary of Discriminative Training Method of Large-scale Feature Set", Journal of Harbin University of Science and Technology, vol. 19, no. 4, **(2014)**, pp. 100-105.
[20] G. Navarro and M. Rafiinot, "Fast and flexible string matching by combining bit-parallelism and suffix automata", ACM Journal of Experimental Algorithmics (JEA), vol. 5, no. 4, **(2000)**.
[21] L. R. Dice, "Measures of the amount of ecologic association between species", Ecology, vol. 26, **(1945)**, pp. 297-302.
[22] M. Kitamura and Y. Matsumoto, "Automatic Extraction of Word Sequence Correspondences in Parallel Corpora", In Proceedings of the Fourth Workshop on Very Large Corpora, Copenhagen, Denmark, pp. 79-87.
[23] I. Dagan, L. Lee and F. C. N. Pereira, "Similarity-based methods for word sense disambiguation", In 35th Annual Meeting of the ACL. Somerset, New Jersey: Association for Computational Linguistics, **(1997)**, pp. 56-63.
[24] R. I. Griffiths, A. S. Whiteley, A. G. O'Donnell and M. J. Bailey, "Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition", Appl Environ Microbiol, vol. 66, **(2000)**, pp. 5488-5491.
[25] A. Maedche and S. Staab, "Ontology learning", In S. Staab and R. Studer, editors, Handbook on Ontologies, Springer, **(2004)**, pp. 173-189.
[26] X. Xiusan, "Physical entropy, information entropy and its evolution equation", Science in China, vol. 31, no. 1, **(2001)**.
[27] B. H. Weber, D. J. Depew and J. D. Smith, "eds. Entropy, Information and Evolution", Cambridge: The MIT Press, **(1988)**.
[28] J. N. Kapur and H. K. Kesavan, "Entropy Optimization Principle with Application", San Diego: Academic Press, **(1992)**.
[29] A. Lakhina, M. Crovella and C. Diot, "Mining anomalies using traffic feature distributions", In: Proceeding of the ACM SIGCOMM. Philadelphia: ACM Press, **(2005)**, pp. 217-228.
[30] H. Bing, H. Xin and Z. X. Zhong, "Rough Entropy Based on Generalized Rough Sets Covering Reduction", Journal of Software, vol. 15, no. 2, **(2004)**, pp. 215-220.
[31] W. Hong and G. Zhenghu, "Algorithm Based on Entropy for Finding Critical Traffic Matrices", Journal of Software, vol. 20, no. 5, **(2009)**, pp. 1377-1383.
[32] http://oaei.ontologymatching.org/2014/food/
[33] http://aims.fao.org/website/About-AGROVOC/sub /
[34] http://agclass.nal.usda.gov/agt/download.shtml/
[35] http://oaei.ontologymatching.org/2014/results/FAO-results.Html