http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 8 (8): 1242-1248, 2009
ISSN 1812-5638

An Intelligent Topic Map-Based Approach to Detecting and Resolving Conflicts for Multi-Resource Knowledge Fusion

Huimin Lu and Boqin Feng
School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, 710049, China

Abstract: In this study, we propose a novel concept of intelligent topic map, which embodies the multi-level, multi-granularity and inherent relevant characteristics of knowledge and realizes knowledge reasoning. With the intelligent topic map as infrastructure, we design a specific ontology fusion process for multi-resource knowledge fusion. Also, we define the taxonomy of merging conflicts which occur during the process of intelligent topic maps merging. We define and classify merging conflicts into data-level conflicts, structure-level conflicts, rule-level conflicts and temporary-level conflicts. We propose the detection and resolution schemes for each merging conflict. Additionally, we implement the multi-resource knowledge fusion conflicts detection and resolution system based on rules. The experimental results show that our method can correctly detect and resolve the conflicts in topic maps merging and it is helpful to improve the quality of multi-resource knowledge fusion.

Key words: Knowledge fusion, intelligent topic map, conflicts detection and resolution

INTRODUCTION

Massive amounts of knowledge which are often geographically distributed and owned by different organizations are being mined. In order for the local knowledge to be reused, they first need to be fused or aligned to one another to produce a single integrated and reconciled the global knowledge that deals with a larger domain of interest. Knowledge fusion can establish a uniform view of distributed heterogeneous knowledge resources for users to provide transparent inquiry and knowledge navigation. In the process of merging, conflicts can be caused by many reasons, like differences of the people's understanding, the marking of knowledge resources and the constructing of knowledge organization. These conflicts can cause information redundancies, contradictions and mistakes and lead to inconsistencies in knowledge fusion. Conflicts detection and resolution is a key component of any knowledge fusion strategy.

In this study, with a new concept of intelligent topic map as infrastructure, we propose a novel approach to merging topic maps. Conflicts are defined and classified in abstract level and described in the formal specification by analyzing the process of merging two topic maps into a new one. After analyzing the proportion of the conflict elements in topic map documents and semantic meaning of the conflict elements, conflicts elimination strategies are drawn up according to the consistency standard and principle. For our experiments, we used a part of topic maps, which are developed from extraction and conceptualization of computer network knowledge.

Many researchers have done a lot of work in multi-resource knowledge fusion field, a few methods for merging between ontologies were proposed, such as Anchor-PROMPT (Noy and Musen, 2001), Chimerae (McGuinness *et al*., 2000), FCA-Merge (Stumme and Madche, 2001), merging models based on given correspondences (Pottinger and Bernstein, 2003). Durusau *et al.* (2006) and Garshol and Moore (2006) described how to merge between entities of topic maps to produce a topic map. However, the method based on the equality of entities must have equal base names and scopes or equal subject identities, which cannot merge between entities that have different structures but have semantic correspondences. Kim *et al.* (2007) proposed a multi-strategic matching and merging approach to find correspondences between ontologies based on the syntactic or semantic characteristics and constraints of the topic maps. The general steps of topic maps merging and their functions were summarized by Lu *et al.* (2009).

Merging entities is a complicated task because it requires finding semantic correspondences between two entities, detecting the various merging conflicts, resolving the detected conflicts, satisfying the merging requirements and generating the duplicate free entities,

Corresponding Author: Huimin Lu, School of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an, 710049, China

etc. (Chung and Kim, 2007). A few methods for detecting and resolving conflicts on knowledge fusion were proposed by Lu and Zhang (2008), Ma and Jin (2007), Wu and Ma (2007) and Gu *et al.* (2004). In the process of merging, conflicts must be detected and resolved to improve the quality of merging results. They do not consider defining merge conflicts problems and providing solutions to those problems.

INTELLIGENT TOPIC MAP

Topic Map is an ISO standard (ISO/IEC 13250) (ISO/IEC JTC 1/SC34 N323, 2002; ISO/IEC, 2008) that describes knowledge structures and associates them with information resources. The structure of topic map is shown in Fig. 1. While, it is possible to represent immensely complex structures using topic maps, the basic concepts of the model-Topics, Associations and Occurrences (TAO) -are easily grasped (Pepper, 2000). Topics define the concepts. Associations define the relationships between the topics and could represent arbitrary number of roles among arbitrary number of topics. The association type for the relationships might be subclass of, part of, instance of, property of et al. they are themselves regarded as topics. The association types make it possible to group together the set of topics that have the same relationship to any given topic and provide intuitive and user-friendly interfaces for navigating large pools of information. Occurrences link the information resources with topics. Topic maps absorb the ideas contained in the semantic web, which established a semantic web above the resource level. The semantic organization and joining between the physical resource entities and the abstract concepts are implemented. But the conventional topic map based on a TAO structure only supports topic navigation and cannot reflect the relevance of knowledge elements. It is difficult to achieve the expression of multi-level, multi-granularity and inner relevant characteristics of the knowledge.

In our framework of the intelligent topic map (Lu *et al.*, 2008), we define a clustering level above the topic level. Furthermore, knowledge element level is inserted above the resource level. Extended topic map structure is shown in Fig. 2.

Each cluster contains several closely related topics. The clustering level provides the effective navigation and browsing mechanism for the users. Knowledge element level allows users to access more detailed knowledge information. Intelligent topic map organizes knowledge from four levels: cluster level, topic level, knowledge element level and resource level and constructs multi-granularity knowledge representation architecture which includes clusters, topics, knowledge elements, associations and occurrences.

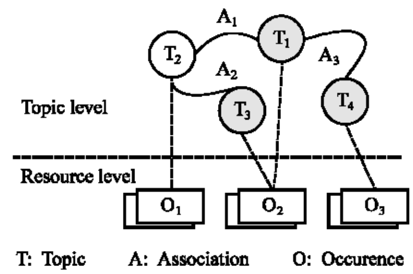

Fig. 1: The structure of topic map

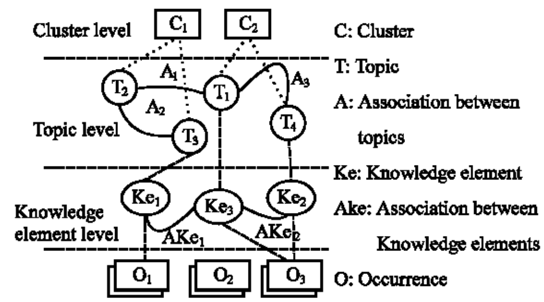

Fig. 2: The structure of extended topic map

Conventional topic map is a graphical index but lack of knowledge reasoning abilities. We established corresponding logical reasoning rules and grammar and then realized knowledge reasoning. Knowledge reasoning mainly includes relationship type reasoning, association reasoning, knowledge architecture reasoning and order reasoning etc.

MERGING PROCESS

Merging between intelligent topic maps describes the process of integrating two local intelligent topic maps into a global intelligent topic map, which is divided into two parts: First, how the similarity of a pair of topics or knowledge elements may be computed and second, how the local intelligent topic maps are merged according to rules. Merging process is shown in Fig. 3.

Similarity computation: Similarity computation is the prerequisite and basis for intelligent topic maps merging. We propose the similarity measure method based on comprehensive information theory. The process used in the similarity algorithm consists of syntactic matching, semantic matching and pragmatic matching (Lu *et al.*, 2009). Syntactic matching is used to compute the

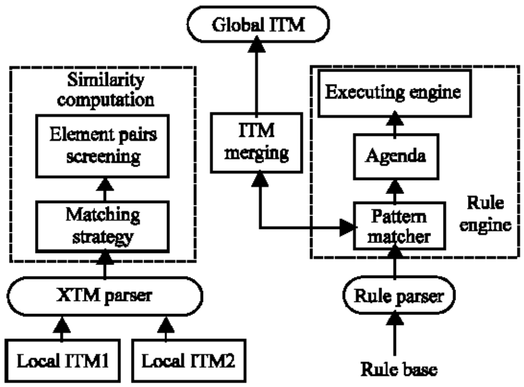

Fig. 3: The merging process of intelligent topic map

syntactic similarity by analyzing the character composition of topics or knowledge elements. During the semantic matching period, the algorithm analyzes the static semantic similarity with aspect to synonyms. Pragmatic matching computes the dynamic semantic similarity, which solves the problem of polysemy. That is, the same word may have different meanings in different contexts.

Mergence: We propose the method of ITM merging based on rule engine. The merging rules are described by rule description language based on topic map. The rule descriptions are saved in the rule documents and the documents are loaded and parsed by rule engine.

TAXONOMY OF MERGING CONFLICTS

We define the taxonomy of merging conflicts which categorizes the conflicts as data-level conflicts, structure-level conflicts, rule-level conflicts and temporary-level conflicts. The taxonomy of merging conflicts is depicted in Fig. 4.

Data-level conflict: It indicates that the same concepts produce the conflicts due to different perceptivity. It is divided into naming conflicts and ID conflicts. For example, let two topics, T_a and T_b, which name or ID have the same values and semantic correspondences, they should be integrated into a unified name or ID.

Structure-level conflict: It indicates that the same concept group produces the conflicts by different logical structure expression. It is divided into hierarchical structure conflicts and property conflicts. For example, let us consider ITM_a that has simple conceptualization in which Physicist topic is specialized with Galileo galilei,

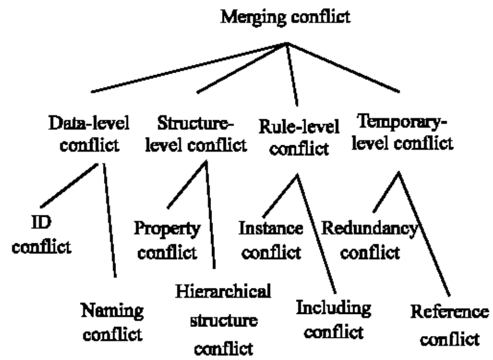

Fig. 4: The taxonomy of merging conflicts

Albert einstein, Xueshen qian, without periodic classification and ITM_b has complex conceptualization in which the Physicist topic is specialized as Ancient physicist, Medieval physicist and Modern physicist by a periodic viewpoint. These ITM_a and ITM_b have structure-level conflicts.

Rule-level conflict: It indicates that the relationship-based conflicts are derived by reasoning rules. It is divided into instance conflicts and including conflicts. For example, let four topics, T_a, T_b, T_c and T_d. If T_c is instance of T_a and T_c is also instance of T_b, but T_a is not related to T_b, this case is instance conflicts. If T_c is subclass of T_a, T_c is also subclass of T_b and T_d is instance of T_c, but T_a is not related to T_b, this case is including conflicts.

Temporary-level conflict: It indicates that it will be arising when the merged topic map has inconsistency in merging. It is divided into redundancy conflicts and reference conflicts.

DETECTION AND RESOLUTION

According to the proportion of the conflict elements in intelligent topic map documents and the semantic meanings of the conflict elements, we designed a conflicts detection and resolution mechanism for intelligent topic maps merging.

Redundancy conflict: Given an intelligent topic map ITM and topics of T_a and T_b, $\forall T_a$, T_b $\in$ITM. We examine SIM_{topic}, which is the similarity value of topics, whether it is 1 or not:

- If $SIM_{topic}(T_a, T_b) = 1$, their conflict is a complete topic redundancy conflict
- If $SIM_{topic}(T_a, T_b) < 1 \wedge T_a.ID = T_b.ID$, their conflict is a partial topic redundancy conflict

When topic redundancy conflict is detected, it can be resolved as follows:

- If it is complete conflict, one of the topics is deleted
- If it is partial conflict, the merged topic is the union of two topics sub-elements

ID conflict: Given an intelligent topic map ITM and topics of T_a and T_b, $\forall T_a$, $T_b \in$ITM. We examine SIM_{ID}, which is the similarity value of topic names, whether it is 1 or not.

If SIM_{ID} (T_a, T_b) = 1, their conflict is a topic ID conflict.

When topic ID conflict is detected, it can be resolved as follows:

A new unique ID is created for one of topics and the reference of the topic is modified.

Name conflict: Given an intelligent topic map ITM and topics of T_a and T_b, $\forall T_a$, $T_b \in$ITM. We examine SIM_{name}, which is the similarity value of topic names, whether it is 1 or not:

- If SIM_{name} (T_a, T_b) = 1, their conflict is a complete name conflict
- If SIM_{name} (T_a, T_b)<1, a name is substring of another name, such as T_a. Name $\subset$ T_b. Name or T_a. Name $\supset$ T_b. Name, their conflict is a partial name conflict

When topic name conflict is detected, it can be resolved as follows:

- If it is complete conflict, one of the topic names is deleted
- If it is a partial conflict, the name of the merged topic is determined with a larger topic name that includes another topic name

Reference conflict: Given an intelligent topic map ITM and association, TT_a, $\forall TT_a \in$ITM,

If ID = TT_a.member(i).topicRef $\wedge \neg\exists T_a \in$ITM ($T_a$.ID = ID), their conflict is reference conflict.

When reference conflict is detected, it can be resolved as follows:

The quoted topics do not exist in current ITM, the association is deleted.

Rule conflict: Given an intelligent topic map ITM and associations, T_a, T_b, T_c, T_d,$\forall TT_a$, TT_b, TT_c, $TT_d \in$ITM

- If $T_c \in$ instanceOf(T_a), $T_c \in$ instanceOf (T_b) and T_a disjoint-with T_b, their conflict is instance conflict
- If $T_c \in$ subclassOf(T_a), $T_c \in$ subclassOf (T_b), $T_d \in$ instanceOf(T_c) and T_a disjoint-with T_b, their conflict is including conflict

When rule conflict is detected, it can be resolved as follows:

- If it is instance conflict, disjoint-with association is deleted
- If it is including conflict, disjoint-with association is deleted

Temporary conflict: We can detect temporary conflicts through examining whether properties, parent topics, association types, or role types referenced by the merged topic are defined in the merged topic map or not (Chung and Kim, 2007). Given a merged topic T_c,

- T_c.OccType$_k$ $\notin$ T_o: {OccType$_i$|1 $\le$i$\le$n}(Undefined property type conflict)
- T_c.AssocType$_k$ $\notin$ R_a: {AssocType$_i$| 1$\le$i$\le$m} (Undefined association type conflict)
- T_c.RoleType$_k$ $\notin$ T_r: {RoleType$_i$|1 $\le$i$\le$l}(Undefined role type conflict)
- T_c.TopicType$_k$$\notin$ T_c: {TopicType$_i$|1$\le$i$\le$p}(Invalid reference conflict)

We can detect property type conflicts, association type conflicts and role type conflicts through examining whether all property types of T_c already exist in the merged topic map and detect invalid references through examining whether referenced topics are already created or not.

Detection and resolution mechanism of knowledge element conflicts are same as the above strategies.

SYSTEM IMPLEMENTATION

In this study, we put forward the mechanism of conflict detection and resolution based on drools rule engine. It is shown in Fig. 5. The main part of the system is rules creation and rule engine module.

Rules creation: By analyzing the elements in intelligent topic maps documents, the possibility of conflicts is determined, models of conflict elements are established and conflict detection rules are described by the production rules. The concept of conflict point is brought forward in order to store conflict elements. The rules are written based on drools' rules file syntax. Detecting and resolving rule files for each conflict type are established. An example is depicted as follows:

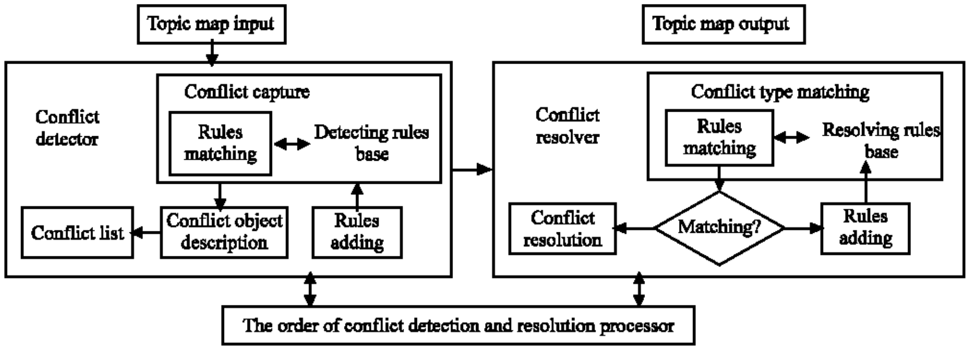

Fig. 5: The mechanism of conflict detection and resolution based on rules

```
rule "detection variantNamePair1"
when
  variant Name Pair: Variant Name Pair (base Name: base Name, variant 1: variant 1, variant 2: variant 2);
  variant Name1: Variant Name (resourceRef1: resource Ref, resource Data1: resource Data! = null) from variant Name Pair. variant Name 1;
  variant Name 2: Variant Name (resource Ref 2: resource Ref, resource Data 2: resource Data! = null) from variant Name Pair. variant Name 2;
then
  variant Conflict Point vcp = new variant Conflict Point (String Compare. compare (resource Data 1, resource Data 2),base Name,variant 1,variant 2);
  insert (vcp);
end
```

Rule engine: A rule engine module is designed, in order to establish the class object knowledge base. The realization of this method is depicted as follows:

```
KnowledgeBuilder kbuilder = Knowledge Builder Factory. new Knowledge Builder ();
kbuilder.add(ResourceFactory.newClassPathResource(rulepath), ResourceType.DRL);
KnowledgeBuilderErrors errors = kbuilder.getErrors();
if (errors.size() > 0)
{
  for (KnowledgeBuilderError error: errors)
{
  System.err.println(error);
  }
throw new IllegalArgumentException("Could not parse knowledge.");
}
KnowledgeBase kbase = KnowledgeBaseFactory.newKnowledgeBase();
kbase.addKnowledgePackages(kbuilder.getKnowledgePackages());
KnowledgeBase kbase = RuleEngine.readKnowledgeBase("rulePath");
StatefulKnowledgeSession ksession = kbase.new Stateful Knowledge Session();
ksession.setGlobal("identifier", value);
ksession.insert("Object");
ksession.fireAllRules();
```

RESULTS AND DISCUSSION

To evaluate the quality of our approach, we need many topic maps constructed by different domain experts of the same knowledge domain. We applied our method to a part of the knowledge domain of computer network. Table 1 shows the characteristics of experimental data.

Table 1: The characteristics of experimental data

Element name	Element No.
topic	50
Knowledge element	93
Association	54
Element assoc	99
Topic element assoc	118

Table 2: The experimental results of detecting and resolving conflicts process

	Conflict detection and resolution results						
Content	TC	KeC	ATC	AKeC	ATKeC	RC	CRC
DCN	4	3	5	3	23	101	1
RCN	4	3	5	3	23	101	1

DCN: No. of conflicts detecting, RCN: No. of conflicts resolving, TC: Topic conflict, KeC: Knowledge element conflict, ATC: Associations between topics redundancy conflict, AKeC: Associations between knowledge elements redundancy conflict, ATKeC: Associations between topics and knowledge elements redundancy conflict, RC: Reference conflict, CRC: Contradictory relation conflict

The experimental results shown in Table 2 describe the statistics of detecting and resolving conflicts.

Experimental results indicate the number of detecting and resolving conflicts according to their types. Reference conflicts rate is 72% in all conflicts, it is because some topics and knowledge elements were deleted in experimental data. The experimental results indicate that conflicts could be correctly detected and resolved in intelligent topic maps.

We evaluated the effectiveness of conflicts detection and resolution during merging process of partial computer network topic map, data link layer topic map, network layer topic map and physical layer topic map. Experimental data of mergence is shown in Table 3.

The experimental results indicate that conflicts could be detected and resolved during intelligent topic maps merging process. The statistical analysis of conflict types are shown in Fig. 6. Some types of conflict are often

Table 3: The experimental data of mergence

LITM	TN	KeN	ATN	AKeN	ATKeN
PCN	194	217	88	197	232
DLL	100	104	112	113	104
NL	100	201	82	200	112
PL	100	100	100	94	44

LITM: Local intelligent topic map, PCN: Partial computer network, DLL: Data link layer, NL: Network layer, PL: Physical layer, TN: Topic No., KeN: Knowledge element No., ATN: Association No. between topics, AKeN: Association No. between knowledge elements, ATKeN: Association No. between topics and knowledge elements

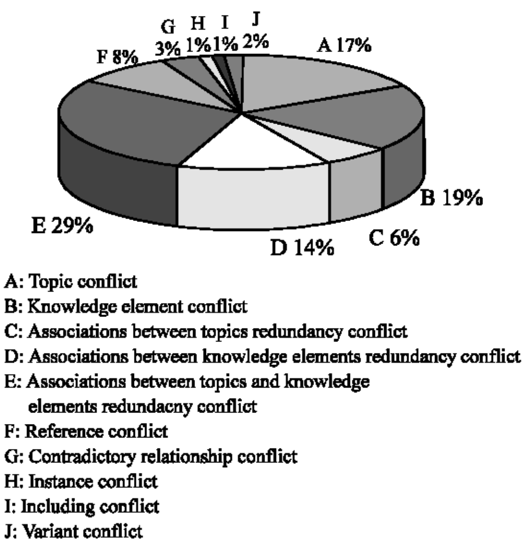

Fig. 6: The statistical results of conflict types

appear, such as topic conflict, knowledge element conflict, relationships between knowledge elements redundancy conflict, relationships between topics and knowledge elements redundancy. But variant conflict, instance conflict and including conflict turn a low frequency. It is due to that document structure and relationship types are relatively simple. Variant conflict involves relatively complicated structure of topic map. Instance conflict and including conflict involve the specific relation types.

From the experiment, we found that our intelligent topic map-based approach to detecting and resolving conflicts shows recall of over 0.9. This means merged topic maps include most of the entities of source topic map without information redundancies, contradictions and mistakes. The reason for the higher recall is that we designed the detailed detecting and resolving rules of conflicts merging. Considering that the huge number of rules may cause system collapse, we classify the rules and carry out inference by means of classification matching. However, the conflicts which are not included in the rules would be handled by the artificial method. Our system-generated automatic detecting and resolving conflicts has superiority over manual detecting and resolving conflicts by an expert, in terms of required time and effort. It is helpful to improve the quality of multi-resource knowledge fusion.

CONCLUSION AND FUTURE WORK

Intelligent topic maps strictly separate store knowledge and associate unstructured information resources, which establish a Semantic Web above the resource level. We realize multi-resource knowledge logical fusion based on intelligent topic map merging. In addition, we put forward a conflicts detection and resolution mechanism based on rules. We hope that, from our intelligent topic map-based approach to detecting and resolving conflicts of multi-resource knowledge fusion, the standards could be made and the real system will be widely deployed in the future.

ACKNOWLEDGMENTS

This research is sponsored by the National High-Tech Research and Development Plan of China under Grant No. 2008AA01Z131; The National Natural Science Foundation of China under Grant No. 60803162. This work is also partially supported by National High-Tech Research and Development Plan of China under Grant No. 2008AA01Z136.

REFERENCES

Chung, H.S. and J.M. Kim, 2007. Conflict detection and resolution in merging of topic maps. Proceedings of the International Conference on Convergence Information Technology, Nov. 21-23, Gyongiu, Korea, pp: 907-912.

Durusau, P., S. Newcomb and R. Barta, 2006. Topic maps reference model. ISO/IEC JTC1/SC34, Version 6.0. http://www.isotopicmaps.org/tmrm.

Garshol, L.M. and G. Moore, 2006. Topic maps data model. ISO/IEC JTC1/SC34. http://www.isotopicmaps.org/sam/sam-model.

Gu, N., J. Xu, X. Wu and J. Yang, 2004. A solution to semantic conflicts in collaborative editing of design documents. Proceedings of the 8th International Conference on Computer Supported Cooperative Work in Design, May 26-28, Xiamen, China, pp: 179-185.

ISO/IEC JTC 1/SC34 N323, 2002. Guide to the topic map standards. International Organization for Standardization. http://www1.y12.doe.gov/capabilities/sgml/sc34/document/0323.htm.

ISO/IEC, 2008. Information technology topic maps part 2: Data model. http://www.isotopicmaps.org/sam/sam-model/data-model.pdf.

Kim, J.M., H. Shin and H.J. Kim, 2007. Schema and constraints-based matching and merging of topic maps. Inform. Process. Manage., 43: 930-945.

Lu, Y. and G. Zhang, 2008. A DL-based approach for eliminating concept restriction conflicts in ontology merging. Proceedings of the 2008 IFIP International Conference on Network and Parallel Computing, Oct. 18-21, IEEE Computer Society, Washington, DC., USA., pp: 446-451.

Lu, H., B. Feng, Y. Zhao, Q. Zheng and J. Liu, 2008. A new model for distributed knowledge organization management. Proceedings of the 7th International Conference on Grid and Cooperative Computing, Oct. 24-26, Shenzhen, China, pp: 261-265.

Lu, H., B. Feng, Y. Zhao, Q. Zheng and J. Liu, 2009. Distributed knowledge fusion based on extended topic maps. J. Jilin Univ. (Sci. Edn.), 47: 543-547.

Ma, Y. and B. Jin, 2007. A combination approach to tackling semantic conflicts based on RDF model. Proceedings of the 3rd International Conference on Semantics, Knowledge and Grid, Oct. 29-31, IEEE Computer Society, Washington, DC., USA., pp: 38-43.

McGuinness, D.L., R. Fikes, J. Rice and S. Wilder, 2000. The chimaera ontology environment. Proceedings of the 17th International Conference on Artificial Intelligence and 12th Conference on Innovative Applications of Artificial Intelligence, Jul. 30-Aug. 3, AAAI Press/ The MIT Press, pp: 1123-1124.

Noy, N.F. and M.A. Musen, 2001. Anchor-PROMPT: Using non-local context for semantic matching. Proceedings of the IJCAI 2001 Workshop on Ontologies and Information Sharing, (WOIS'2001), Seattle, WA., USA., pp: 63-70.

Pepper, S., 2000. The TAO of topic maps. http://www.ontopia.net/topicmaps/materials/tao.html.

Pottinger, R.A. and P.A. Bernstein, 2003. Merging models based on given correspondences. Proceedings of the 29th International Conference on Very Large Data Base, (ICVLDB'2003), Berlin, Germany pp: 862-873.

Stumme, G. and A. Madche, 2001. FCA-merge: Bottom-up merging of ontologies. Proceedings of 7th International Joint Conference on Artificial Intelligence, (IJCAI 2001), Lawrence Erlbaum Associates Ltd., USA., pp: 225-234.

Wu, K. and Y. Ma, 2007. An approach to tackling semantic conflicts based on ontology. Comput. Eng. Appl., 43: 182-185.