



An Advanced ontology based automatic approach in improving the similarity by means of combining the sub-graphs between the information

K. L. Sumathy^{1,*}, M. Chidambaram²

¹Research Center, Bharathiyar University, Coimbatore, India

²Rajah Serfoji Govt. College, Thanjavur, Tamil Nadu, India

ARTICLE INFO

Article history:

Received 22 June 2016

Received in revised form

5 August 2016

Accepted 5 August 2016

Keywords:

Information retrieval

Manufacturing systems

Knowledge modeling

Semantic similarity and ontology

ABSTRACT

In this research an advanced and automatic approach in finding the new products which are similar to the previous one that results in a rapid experience in production of design of manufacturing procedure of the new product has been proposed. The proposed work is based on advanced ontology based semantic model which computes similarity between the sub-graphs in an effective manner. It builds a hierarchical structure by means of new similarity index that forms by overlapping sub-graph of existing two product concepts. By means of stored data the similarity measurement is calculated by matching the similar characteristics with the new one that helps in discovering knowledge. The examined result with the real-time data shows minimum computation cost along with high processing speed in similarity according to the global environment. Thus proves the proposed scheme is far better than other existing similarity approaches.

© 2016 The Authors. Published by IASE. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Generally information is the base of work in any simple business or dealing with a huge manufacturing industry. The simple question arises on the importance of information retrieval in running a business of all kinds. And the answer is information retrieval determines the time consuming and computation costs which results in the business hierarchy in a better way. According to the current industrial market there are several major drawbacks in handling the concepts of storing and reusing the manufacturing knowledge which are available in the companies.

To handle these issues and design a prominent solution in achieving the information retrieval, the best practicing method is ontology. Now-a-days information was handled in web based scenario and the question is how it is going to be handled in better way. Ontology can be built by certain representation of concepts which belongs to the domain. These representations are based on the impact of time, action, physical objects and various other factors. According to which ontology shares structural

similarities and enables the process in an effective manner. In the real time environment, when an organization building a new factor in their business aspect, they need to learn the whole terminology even in which most of the time they will not have adequate knowledge on manufacturing techniques.

This can be fulfilled by developing a best approach which will make it possible to perform the similarity of existing product with the new product in an automatic manner. By means of this novel approach, huge time will be saved with the concept of reusing the information that is already available in the organization's knowledge management. In addition, maintaining of huge information in an organization without the knowledge of existing data will result in memory consumption and drawbacks pertaining to maintenance in the overall level.

2. Related works

Due to the growth of various manufacturing industries and technological growth, there is an emergence of concept deals with automatic approach in finding the information of a management in an organization. Various research works are carried in developing an automatic analysis of available information and reusing it in a simple manner. According to (Bruno, 2015a), it is not simple to extract the definite manufacturing knowledge in their organizations. By the analysis of work on (Lowe et al., 2004) the designer took 20% of

* Corresponding Author.

Email Addresses: sumathyphd@yahoo.com (K.L. Sumathy),

chidsuba@gmail.com (M. Chidambaram)

<http://dx.doi.org/10.21833/ijaas.2016.08.001>

2313-626X/© 2016 The Authors. Published by IASE.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

her time in searching and analyzing the appropriate information available in an organization.

In paper (Papakostas et al., 2010) it is shown by means of personally stored information, 40 % of required information can be gathered. According to the earlier employees worked on the enterprises, they had wide knowledge on documentation that will be helpful in future. The study on (Huang and Diao, 2008) proves that those documentations were difficult to understand and usage was not very prominent for the employees who will work in future, due to the documentation format as well as their standards. In the papers (Denkena et al., 2007), due to the change business standards, about 50% of information was not presented in information system. Moreover the concept of information retrieval and reuse are not applicable for the systems, because they are only documented but not managed.

The evolving of semantic technology and design of knowledge management systems are helpful in realizing the details of companies in a minimum time consumption and reusing it effectively were shown in (Fortineau et al., 2012; Bruno et al., 2014). On this way ontology enables the possibility of combining information based on its abstraction levels and method of improvement in capturing as well as reusing (Chang et al., 2006). According to SMEs (Efthymiou et al., 2015), the knowledge management lack in performing the classification of past projects which take more time in identifying the similarity between new and past projects available in the memory.

From the work on (Liao et al., 1998; Bouzid et al., 2013) they show the similarities only on numerical attributes, otherwise on string attributes it based on edit distance functions. These works do not consider the meaning of the word and that is considered as a major drawback on these schemes. The Gene Ontology context (Teng et al., 2013) deals with similarity on bioinformatics which process the similarity between the genes. But it's not an effective model in achieving the efficiency in terms of medical diagnosis. On the paper (Groover, 2007; Bruno, 2015b), Groover discussed about the Fundamentals of modern manufacturing: materials processes, and systems in forming ontology for manufacturing process.

Several frameworks were developed by ontology using Jena, an open source Java framework for semantic applications. These tools were developed using APIs for RDF with SPARQL queries and supports Protégé ontology editor. In croft research work, the author demonstrates the semantic relationship by means of semantic network (Agosti and Crestani, 1995; Rau, 1987). According to Simone Ponzetto and Michael Strube they created a graph by representing inheritance between the words in Word Net (Knappe et al., 2003; Lopez et al., 2010). But these works require more time and possibility of mismatching, because there is no adequate information in analyzing the meaning of the word used as query.

3. Problem definition

In this section, we discussed about the various problem of manufacturing enterprises facing in the industrial market. In an organization the products were described by its product model which contains factors like technical data in order to produce the product in an appropriate manner. In detail the process model defines various data and information which gives important factors like processing time and requirements of the products. In other aspect the operations like dispensation, assembling, storing, managing, transporting and inspecting. The issues are on the basis of information on the management system and it may vary according to the product.

On retrieving the information, there is to be more similarity of new products with the existing one that will be more efficient than others. Here finding the similarity between the existing and new one took huge time consumption and computation cost. The important drawback is the existing system fails to consider the meaning of the word. It may lead to the possibility of finding similarity between the unrelated products. In addition, an information management fails to show the similarity between the products which may lead to load more information and that is already available in the management system. This increase data duplication and occupies more amount of memory space that also increase the time during the information retrieval process.

4. Proposed system

In order to overcome the issues, we involved in designing a supporting tool which indentifies automatically best product that is similar to the existing one which maximizes the speed of manufacturing new products. The mechanism of the proposed system is, by means of similarity index existing product is overlapped with new proposed which has maximum similarity. The ontology structure is formed on these bases by understanding the meaning of the word. In these manner sub-graphs were created to know the best knowledge in understanding the information management system of an organization. Before creating our proposed frame, it is important to understand the concept of the proposed system. Fig. 1 shows the overall mechanism of the proposed system.

As in Fig. 1, the ontology system separates system into function knowledge and manufacturing knowledge. The functional knowledge gives details about the product functionalities which is helpful to collect the information related to manufacturing.

By means of that, ontology creates mapping structure between the data and semantics tree structure. Based on the indentified functions, a link between the sustainable manufacturing and functions of manufacturing processes is created.

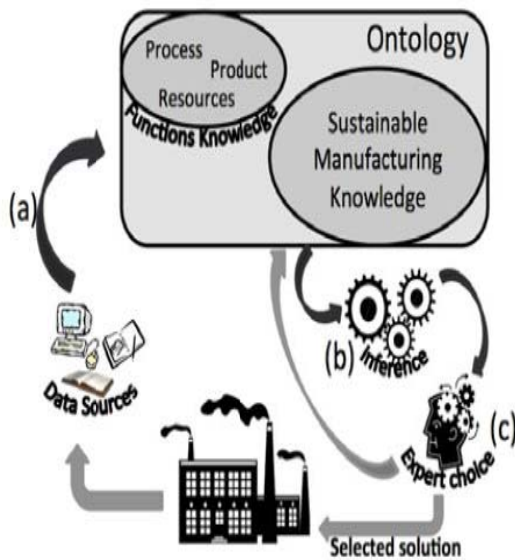


Fig. 1: The ontology and path of executing automatic tasks

The intension of this system is to retrieve the useful details of the product along with manufacturing-related information for our proposing framework. So the framework easily identified and the existing product with the new product will be automatically in the information available in the organization management system.

Before designing a framework it is crucial to know about ontology and how mapping can be done in an effective manner. Generally ontology creates semantic mapping in order to ensure their interoperability between them. Ontology is a best solution for semantic heterogeneity problems dealing with information integration and sharing. Traditionally ontology mapping techniques use string-based methods and linguistic-based methods. Here it's formed mapping between the relationships among the entities. On ontology mapping a vital role is played by structural information. The best entity will form mapping by means of its meaning.

Fig. 2 shows the entity with two different labels such as Proceeding and Proc with same meaning. It is not semantically very similar, but they belong to ontology by means of its neighboring entities such as reference, book, monograph and collections. Before building the framework, some steps to be followed in ontology development which is shown in Fig. 3.

Step 1: Before forming ontology one should know the important terms and concepts by which the similarity is achieved in an exact manner that is explained in Simple Knowledge Organization System (SKOS) (Kolte et al., 2009).

Step 2: The Web Ontology Language will have information on (OWL) (Badder et al., 2003) making classes and their proprieties in a proper way, by which the mapping between them based on the relationship of the product can be made. It also satisfies the logical rule in which the mapping is to be computed.

Step3: Semantic Web Rule Language (SWRL) (ElMaraghy, 2005) which provides the rules and reasoning factors in ontology.

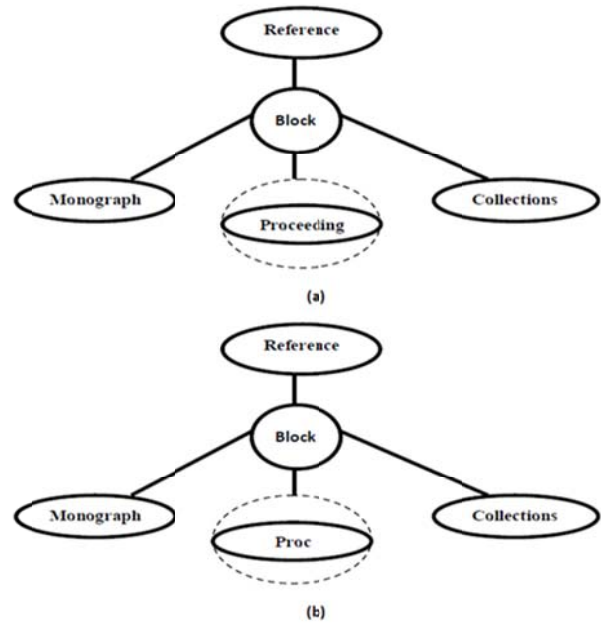


Fig. 2: Entity with same meaning

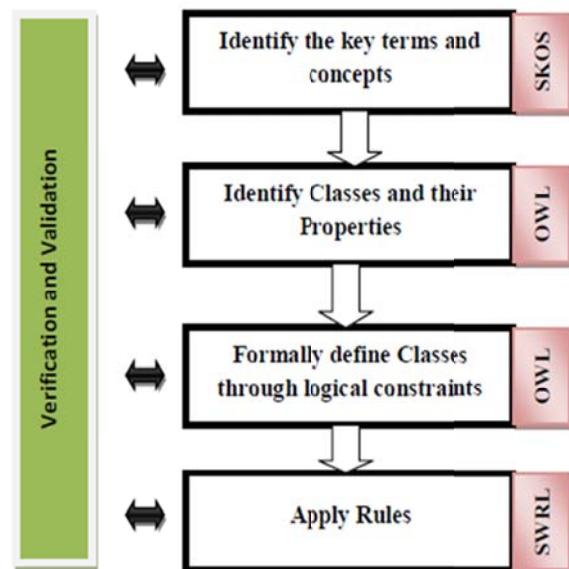


Fig. 3: Steps of ontology development process

5. Framework for semantic similarity computation

This section gives the details about the design proposed framework intended to show the best semantic similarity between the existing and new product. In which, the initial work is in understanding the concept of manufacturing ontology process. It contains of two factors such as processing operation and assembly operations. The processing operation gives the details of result and working of products on its various stage in which sub-graph were created using ontology. In the same manner the assembly operation shows the details of assembling the order of information's helpful for manufacturing the products. Here also a sub-graph is created based ontology, if necessary overlapping sub-graphs can be done to obtain the final result (Fig. 4).

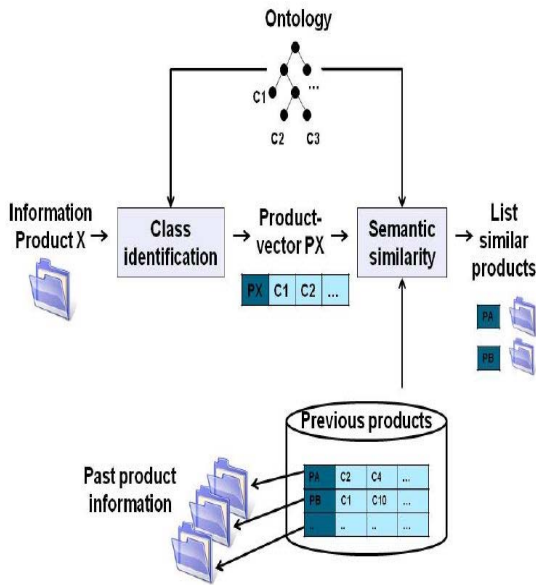


Fig. 4: Framework for semantic similarity

Fig. 4 shows the mechanism of proposed framework in finding the similarity between the two products. By which ontology creates tree and sub-tree matching the products similarity according to the given query. The new product to be manufactured is considered as product X. Next, the product undergoes process of identification automatically with class identification and match with the past product available in the information system. The semantic similarity is achieved with the product X by means of the achieved maximum similarity. The product X completes the manufacturing after several factors like grinding, shaping, cleaning, processing, annealing, surface processing and coating.

6. Sub-graph overlapping process

Fig. 5 shows the possibility of process which would be undergoing certain processes before completing its manufacturing stage. Various tree structures are formed with certain connectivity with one another. As discussed above, various factors such as grinding, shaping, cleaning, processing, annealing, surface processing and coating were assume with $c_1, c_2, c_3, \dots, c_N$ respectively. Based on these relationships the tree structure is formed. Considering a new product to design, initially an organization has to check with the old information management. Based on the characteristic, the ontology creates a sub graph on the particular domain. From which the factors matching with new and existing data overlap each other in a prominent manner as shown in the figure. Thus according to the process, various sub-graphs such as G1, G2 and G3 are formed based on the product similarity achieved. Then by means of maximum similarity, union between the sub-graphs was formed automatically. The maximum similarity between the two sub-graphs formed using the expression $G1 \cap G2$.

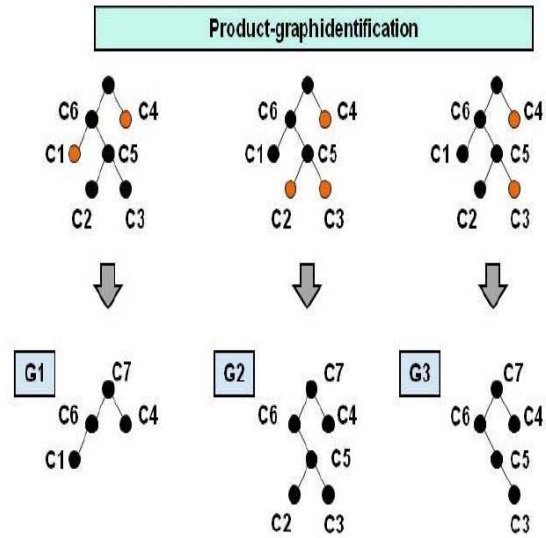


Fig. 5: Sub-graph overlapping system

7. Performance analysis

The semantic similarity is achieved by calculation based on the information content obtained and processed. By means of the Fig. 3 similarities are calculated using the expression (Eq. 1);

$$sim(c_1, c_2) = -\log \frac{len(c_1, c_2)}{2 * D} \tag{1}$$

Here, D describes the similarity distance between the overlapped sub-graphs.

The D distance between two categories is calculated by the amount score gained between them. It can be shown as;

- Identity: $D(c_1, c_2) = 0 \Leftrightarrow c_1 = c_2$
- Normalization: $0 \leq D(c_1, c_2) \leq 1$
- Triangle inequality: $D(c_1, c_3) \leq D(c_1, c_2) + D(c_2, c_3)$

Next the information in the content analysis by using the expression (Eq. 2);

$$IC(c) = -\log(p(c)) \tag{2}$$

where IC determines information content, p shows product and c describes the combination of factors involved for manufacturing the product as shown in Fig. 3. The proposed semantic similarity is achieved by (Eqs. 3-5);

$$sim(c_1, c_2) = \max_{c(c_1, c_2)} IC(c) \tag{3}$$

The final result achieved is;

$$sim(c_1, c_2) = \frac{2 * IC(C)}{IC(C_1) + IC(C_2)} \tag{4}$$

$$sim(c_1, c_2) = G1 \cap G2 \tag{5}$$

The above expression is computed based on product X which undergoes manufacturing. Before manufacturing the similarity is calculated with old

information in the organization. By means of the expression the maximum similarity is achieved with the product X. In these various factors involved in manufacturing a product is considered, that is c1, c2, ..., cN. Those are tabulated on the basis of information content (IC) best matching which is shown in Table 1 and 2. Here the calculation based on the sample product X undergoes various factors

such as grinding, shaping, cleaning, processing, annealing, surface processing and coating and the values are tabulated. To analyze the proposed works efficiency the results are shown graphically in Fig. 6 based on the tabulation. In Table 2 it is clear that the new product X is matched with three existing ones and best one with maximum similarity is taken into the consideration.

Table 1: Information content (IC) of each class

Class	D (C)	Log D((C) +1)	Log D((C) +1) + Log(TOT)	IC
C1:Grinding	0	0	0	1
C2: Cleaning	0	0	0	1
C3: Coating	0	0	0	1
C4:Annealing	0	0	0	1
C5: Surface Processing	2	0,49	0,57	0,47
C6: Shaping	4	0,73	0,85	0,18
C7: Processing	8	0,88	1	0

Table 2: Semantic similarities computed between the three sample products.

	P1	P2	P3
P1	1 $(0+0,18+1)/(0+0,18+1+0,47+1+1)=0,323$	$(0+0,18+1)/(0+0,18+1+1)=0,541$	$(0+0,18+1)/(0+0,18+1+0,47+1+1)=0,73$
P2		1 $(0+0,18+1+0,47+1)/(0+0,18+1+0,47+1)=1$	
P3			1

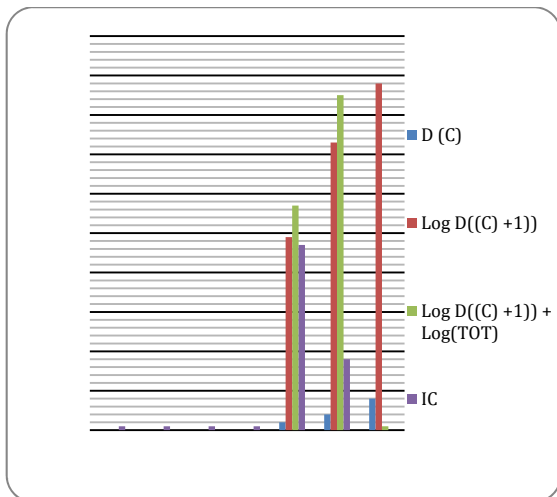


Fig. 6: Graphical representation of the similarity

8. Conclusion

In this paper we discussed about the problem of identifying similarity between the existing and new product. By which a proposed model framework is designed using a modern approach of sub-graph overlapping based on maximum similarity. The information content and ontology formation between the products are achieved automatically, thus proving the effectiveness of our proposed work. The result and work shows that new product manufacturing is built with maximum speed in extracting the information, and forms product graph along with sub-graph proving as a prominent solution then the other existing works in the industry.

References

Agosti M, Melucci M and Crestani F (1995). Automatic authoring and construction of hypermedia for information retrieval. *Multimedia Systems*, 3(1): 15-24.

Badder F, Calvanese D, McGuinness DL, Nardi D and Patel-Schneider PF (2003). *The description logic handbook: Theory, Implementation and Applications*. Cambridge University Press, Cambridge, United Kingdom.

Bouzd S, Cauvet C, Frydman C and Pinaton J (2013). A semantic support to improve the collaborative control of manufacturing processes in industries. In *Computer Supported Cooperative Work in Design (CSCWD)*, 2013 IEEE 17th International Conference on, Whistler, BC, Canada: 433-438.

Bruno G (2015a). Measuring product semantic similarity by exploiting a manufacturing process ontology. In *Industrial Engineering and Systems Management (IESM)*, 2015 International Conference on, Seville, Spain: 1251-1257.

Bruno G (2015b). Semantic organization of product lifecycle information through a modular ontology. *International Journal of Circuits, Systems and Signal Processing*, 9: 16-26.

Bruno G, Antonelli D, Korf R, Lentes J and Zimmermann N (2014). Exploitation of a semantic platform to store and reuse plm knowledge. In *IFIP International Conference on Advances in Production Management Systems*, Springer Berlin Heidelberg, 438: 59-66.

- Chang X, Sahin A and Terpenney J (2008). An ontology-based support for product conceptual design. *Robotics and Computer-Integrated Manufacturing*, 24(6): 755-762.
- Denkena B, Shpitalni M, Kowalski P, Molcho G and Zipori Y (2007). Knowledge management in process planning. *CIRP Annals-Manufacturing Technology*, 56(1): 175-180.
- Efthymiou K, Sipsas K, Mourtzis D and Chryssolouris G (2015). On knowledge reuse for manufacturing systems design and planning: A semantic technology approach. *CIRP Journal of Manufacturing Science and Technology*, 8: 1-11.
- ElMaraghy HA (2005). Flexible and reconfigurable manufacturing systems paradigms. *International Journal of Flexible Manufacturing Systems*, 17(4): 261-276.
- Fortineau V, Paviot T, Louis-Sidney L and Lamouri S (2012). SWRL as a rule language for ontology-based models in power plant design. In *IFIP International Conference on Product Lifecycle Management*. Springer Berlin Heidelberg: 588-597
- Groover MP (2007). *Fundamentals of modern manufacturing: materials processes, and systems*. John Wiley & Sons, New Jersey, USA.
- Huang N and Diao S (2008). Ontology-based enterprise knowledge integration. *Robotics and Computer-Integrated Manufacturing*, 24(4): 562-571.
- Knappe R, Bulskov H and Andreasen T (2003, October). Similarity graphs. In *International Symposium on Methodologies for Intelligent Systems*, Springer Berlin Heidelberg: 668-672
- Kolte SG and Bhirud SG (2009). WordNet: a knowledge source for word sense disambiguation. *International Journal of Recent Trends in Engineering*, 2(4): 213-217.
- Lopez V, Fernández M, Motta E and Stieler N (2012). Poweraqua: Supporting users in querying and exploring the semantic web. *Semantic Web*, 3(3): 249-265.
- Lowe A, McMahon C and Culley S (2004). Information access, storage and use by engineering designers, part 1. *The Journal of the Institution of Engineering Designers*, 30(2): 30-32.
- Papakostas N, Efthymiou K, Chryssolouris G, Stanev S, Ovtcharova J, Schäfer K and Eytan A (2010). Assembly process templates for the automotive industry. In *3rd CIRP Conference on Assembly Technologies and Systems*, Trondheim, Norway: 151-156.
- Ponzetto SP and Strube M (2007). Deriving a large scale taxonomy from Wikipedia. In *AAAI-07, 22nd International Conference on Artificial Intelligence*, Vancouver, British Columbia, Canada,: 1440-1445.
- Rau LF (1987). Knowledge organization and access in a conceptual information system. *Information Processing and Management*, 23(4): 269-283.
- T Liao TW, Zhang Z and Mount CR (1998). Similarity measures for retrieval in case-based reasoning systems. *Applied Artificial Intelligence*, 12(4): 267-288.
- Teng Z, Guo M, Liu X, Dai Q, Wang C and Xuan P (2013). Measuring gene functional similarity based on group-wise comparison of GO terms. *Bioinformatics*, 29(11): 1424-1432.