Automation of Concept Map Generation using Correlation and Stop Word Identification

Ramasamy Kumar

Professor, Department of Information Technology,
Rajalakshmi Engineering College, Chennai, Tamil Nadu, India
Email: kumar.r@rajalakshmi.edu.in

Abstract: *Concept Map is a technique used for visual representation of concepts and their relations. Currently concept maps are constructed manually which can be a complex task and that needs skills otherwise the person may experience difficulties in determining and positioning concepts. A major expectation amongst researchers is on a method that creates concept maps automatically. A tool that assists the user in identifying and placing concepts and their relations in appropriate levels and position, would be highly appreciable in the present scenario. This paper proposes a new method to automate the creation of map from a given sentence or paragraph. This method involves Stop Word Identification, Correlation approaches. Initial results are encouraging and motivates to perform a rigorous research.*

Keyword: *Concept Map; Automation; Concept Word; Keyword; Stop words.*

1. INTRODUCTION

A concept map (CM) is regarded as a knowledge representation tool that has been widely used for organizing and representing knowledge. It includes the following elements; 1. Concepts, usually signified by nouns or noun phrases, 2. Relationship between concepts indicated by a directed line or arc linking two concepts and 3. Labels that represent the relationship among concepts usually a verb or a verb phrase creates a concept-label-concept chain. This meaningful chain can be read as a sentence and this chain is called a proposition (Novak & Canas 2008).

Recent years evidenced a remarkable growth in the usage of CMs in various domains. The most prevalent applications of CMs are reported in Education as CMs are found more effective in facilitating meaningful learning.

In a teaching & learning environment, a CM can be used as a tool that represents the teaching contents, Lesson plan, and knowledge acquired. For most learners especially for beginners it is difficult to build a map for a chosen topic of interest or focused question. A skeleton map provided by an expert can make it easier for the learner to start that process. In personal learning, it is difficult to find experts for specific learning fields. Therefore, a system that behaves like an expert and provides the skeleton of a CM can be very helpful in such situations. The creation con-

cept map is a tedious processes which needs lot of patience. Time taken for creating a map is regarded as one of the most important challenges in Concept Mapping. This motivated researchers across the world to look into possible ways of automating the concept map creation. Concept maps are known as node-arc kind of visual representations of knowledge that show concepts as nodes and the relationships between the concepts as links. This paper introduces a method which is semi-automatic creation of a CM from unstructured text.

The rest of this paper is organized as follows; Section 2 presents the existing works, section 3 covers the proposed technique with algorithm proposed. Section 4 explains the implementation of the technique with the outcomes. Section 5 concludes the article.

2. RELATED WORKS

A concept map created automatically will provide an overview of knowledge from a document or domain chosen. It facilitates the student's understanding of the concepts. Constructing a good concept map from the given text requires correct selection of concept words and identification of meaningful relationships between the concepts. Studies on automatic creation of concept maps have been reported from many years. It is learned that earlier works discussed several challenges such as an inconsistent definition of concept maps, and a lack of an evaluation framework [7]. It is very difficult for us to conclude what a concept map means. Researchers reported the difficulties and challenges in creating and evaluating the CMs. Further it is very difficult for a human expert to say for

certain what a "correct" concept map should look like, hence the need for creation of automatic concept maps using either semi-automatic or automatic approaches is justified.

The automation is experimented with concept maps in two ways. First one is to automate the creation of CM while the second one is automation of assessment processes using CM thus a significant role of automation is used for evaluation of CMs.

Villalon & Calvo [7] regarded the automatic or semi-automatic creation of CMs from documents as concept map mining (CMM). In the earlier, the system identifies and recommends the concepts and relations of a map, and a user has to approve and finish the map. In the later, the system will create the map automatically from the sources.

A plug-in to CmapTools was developed by Hayes [3] which allows the automatic generation of concept maps from ontologies and Graudina and Grundspenkis [4] also described a tool that supported this task. Several works were reported on development of tools that creates maps from text. Richardson *et.al* [5] have developed a tool named Relex that produces concept maps from text.

Cabral and Goncalves [6] describe software that transforms concept maps into SCORM-compliant learning objects, and represented in another XML vocabulary. Automation of a concept map includes many technique [1], few such are Parsing, Summarization, POS Tagging, Correlation, Stemming, Stop word identification, Hidden Markov Model, etc. Each technique is either used separately or combined. Usually a combined technique gives best results.

Novak and Canas [2] stated that an expert skeleton map pave a better foundation for learning further they create a chances by which additional concepts and propositions can be added. In This principle applies regardless of whether the concept maps are generated by humans or by automated means. Similarly, Lahti [8] proposed guided generation of concept maps from open access online knowledge resources such as Wikies. He also implemented a prototype extracting semantic relations from sentences surrounding hyperlinks in the Wikipedia's articles. [9]

Wafula [10] introduce a new automatic method for constructing concept maps. The heuristic applied to extract concepts is term occurrence. A similar principle is applied in extracting relations. Initial results show that sensible concepts and nouns occur more frequently in a given test material. More sensible relations between concepts also occur more frequently in the text. With syntactic analysis and auxiliary ontologies, term occurrence can be seen as a viable approach to constructing fully automatic concept maps.

Karannagoda [11] proposed a method that analysis

the document for automatic creation of concept maps. A rule based approach is used to extract concepts and relationships among them. Sentence level breakdown enables these rules to identify those concepts and relationships. These rules are based on elements in a phase structure tree of a sentence. The concept map is visualized using force directed type graphs which represent concepts by nodes and relationships by edges. Lee *et al.,*[12] reported an automatic concept map creation using a burst analysis approach from a single document [12].

Sun [13] introduced several methods for automated concept mapping between medical databases by using concept maps and ontologies. The use of conceptual context to perform automated concept mapping facilitates the identification of equivalent database concepts. It helped in decreasing the work and costs associated with information management from disparate databases. Similarly Luckie *et al.,* [14] used automated grading of concept maps during the online assessment processes. *Concept Connector*, consisting of a web-based application with automatic scoring of concept maps. It was designed to enable students to visualize their thinking online and receive immediate formative feedback. Similarly CMs are used for learning assessment widely [15]

Nugumanova et al.,[16] reported a novel approach of automatic construction of concept maps for various knowledge domains by a 3-steps algorithm for extraction of key elements of a concept map (nodes and links) from a given collection of domain documents. This work utilizes text mining approach to achieve the goals.

Leake *et.al* [9] reported exploiting the information in concept maps to automatically generate and refine queries to Web search engines, to aid concept mapping. Apart from these some researchers use self-organized maps (SOM) for learning purposes [17].

3. METHODOLOGY

Concept maps are node – arc based graphs which are used to represent the concepts and relations of a given sentence or paragraph or document. Drawing a concept map manually needs a better understanding of the core concepts, their hierarchy and respective relationships. The concept map construction may vary from person to person, but the representation must be similar if not identical Concept map can be automated to get a unique representation of a concept.

Automation of concept map should follow all the principles and rules of the concept mapping process. That is the map derived should give the flow of concept and should also flow the rules such a cross relation, hierarchy. Also an automated map should be able to show the flow of concept in the given text.

Automation of concept map begins with the identification of keywords from the given text. Then from the key words the concept words and relationships for the construction of map. This work uses two techniques such as correlation and stop word identification. These techniques are used to identify the key words, concept words, concept word hierarchy and relationships.

3.1 Key Word Identification

The action words or words which describe the action performed in the sentence are considered as Keywords. These actions can describe the work done thus the topic, or the type of being functions that document is saying about. The keywords of a document can be any form grammar. The most suitable form English grammar is the nouns in the sentence. The nouns can be considered as the keywords because the action of a sentence is centered on the noun present.

The nouns in the sentence can be identified by using the stop word removal technique. Stop word are the words such as conjunctions, adjective and prepositions. Conjunctions, adjective and preposition are considered as stop words because they are used as fillers which will help the sentence get a form.
The software developed will remove the stop words in sentence by using the databases stored with the stop words. A single stop word should be stored in the all tense of grammar. This is done because software just compares each word of given document and removes the stop words stored in the database. The problem of storing all possible tense form of a word can be overcome by the use of stemming algorithm.

3.2 Relationship Identification

There are no specific rule define for a relationship to derived from the document. While using the P.O.S technique we can use the verb as the relationship definer. A verb can used has a relationship because it defines the action done by a sentence. Also while considering the cross reference for a node in the map the verbs will not from a perfect relationship.

A defined relationship for a set of node can't be surely identified. There are no specific rules for what should be a relationship between nodes. Randomly any word can be chosen in the sentence for relationship, which makes it difficult to define the relationship to the automated system.

3.3 Concept Word Identification

The keywords identified are used to identify the possible concept word. The keywords are first paired to form the key pairs. The pairing is done as follows, (K_1, K_2), …., (K_1, K_n), (K_2, K_3), …, (K_2, K_n), …, (K_{n-1}, K_n). After paring the Correlation for each pair is found. Correlation is relation between two words and

probability value lies between -1 and 1. Correlation is found using formula (1) given below:

$$RS(K_i, K_j) = \log_{10} \begin{cases} \dfrac{\frac{n_{i,j}}{\max(n)}}{\dfrac{avg\ d_{ij2}}{\max(avg\_d^2)}} & i \neq j \end{cases} \quad (1)$$

3.4 Concept Map Generation

The concept word pairs found using the correlation is ordered in ascending order. The pairs are checked for continuous occurrence and the level for the map is fixed. The level found may have cross links for the concept map. The cross links are treated with values found with correlation. That is, if a keyword has a link with two other key words, which in turn are linked to each other, the lower most leaf or node's or child's link will be given to the ancestor or node or leaf at the highest level.

- (K_i, K_j) expresses the correlation degree between keyword i and keyword j
- n_{ij} expresses the frequency keyword i and keyword j appear in the same sentence

$$avg\_d_{ij2} = \sum \left(\frac{d_m2}{n_{ij}} \right) \quad i \neq j \quad (2)$$

- Max(n) is the number of Sentence the keyword is present
- Max(Avg_d) is the number words present in the keywords present sentences
- d is the distance between the two keywords

A correlation value between -1 and 0 is not considered because there exist no relation between the two words. Only values between 0 and 1 are considered. The correlation value found is used to draw the concept map. The concept words and correlation are identified using equations (1) and (2). However manual intervention in this work is unavoidable as the accuracy of the map needs to improve.

4. IMPLEMENTATION

The construction of concept map generation was fully developed using java as the UI and MySQL 5.0 as the backend. Also a frame work called the JWizard Component Ver: 1.0 developed by William Ready is used for the GUI. The tool is designed to work semi automatically, i.e. the identification of stop words and verbs have to be approved by the user. The verbs and stop words are stored in the database and used accordingly.

The tool uses a jar dictionary which helps to identify the word given is off meaning. The user either gives a sentence or the whole document. The sentence is taken and preprocessed. Preprocessing does the job

of meaning check and for the occurrence of special character, white spaces.

After preprocessing the stop words are removed from the sentence. Then the verbs are removed leaving only the nouns present in the sentence. Then the nouns are paired to find the concept words.

The correlation is found using the formula which is converted into a simple algorithm which is as follows:

```
initialize the variables
   read c1 //no of sentence
   for i!=c1
      read c2 //no of occurrence
      read h; //no of words in the sentence
   for j!=c2
   read c; // c is the distance between two keyword
   d->d+(c*c);
end
c3->c3+c2;
avg_d->d/c3;
      h ->h+h1;
m->m+c2;
end
den->den+(avg_d/h);
neu->m/c1;
l->neu/den;
r->log(l);.
```

After finding the correlation the map is drawn by eliminating the pair's value with 0 and -1's. Then the pairs are ordered ascending with their values. Java graphics is used to create the symbols and place the pairs in the levels of the map.

5. CONCLUSION

A new method that is semi-automated for concept map generation is designed and implemented. Initial results are encouraging towards the achievement of objectives. However few erroeneous results are encountered during the evaluation of the method. It needs rigorous analysis of the results to improve the efficiency of the method. The method will also involve Summarization, Stemming, P.O.S tagging or parsing in future research works to study the outcomes. Further, the work needs more research to improve the accuracy of the map.

REFERENCES

[1] Zubrinic, K., Kalpic, D., and Milicevic, M. (2012). The automatic creation of concept maps from documents written using morphologically rich languages. Expert Systems with Applications, 39(16), 12709–12718.

[2] J. D. Novak & A. J. Cañas," The Theory Underlying Concept Maps and How To Construct Them", Technical Report IHMC CmapTools 2006-01, Institute for Human and Machine Cognition (IHMC), 2006.

[3] Eskridge, T., & Hayes, P. (2006). "Formalizing the Informal: A Confluence of the Semantic Web and Concept Maps". In: A. J. Cañas & J. D. Novak (Eds.), *Concept Maps: Theory, Methodology, and Technology. Proceedings of the Second International Conference on Concept Mapping*. San José, Costa Rica:Universidad de Costa Rica, pp. 247-254.

[4] Graudina, V., & Grundspenkis, J. (2008), "Concept Map Generation from OWL Ontologies". In A. J. Cañas, P. Reiska, M. Åhlberg & J. D. Novak (Eds.), Concept Mapping: Connecting Educators, Proceedings of the Third International Conference on Concept Mapping, Tallinn, Estonia & Helsinki, Finland: University of Tallinn, pp. 173-180.

[5] Richardson, R., Goertzel, B., and Fox, E. (2006). Automatic Creation and Translation of Concept Maps for Computer Science-related Theses and Dissertations. In: A. J. Cañas & J. D. Novak (Eds.), *Concept Maps: Theory, Methodology, Technology*. Proceedings of the Second International Conference on Concept Mapping. San José, Costa Rica: Universidad de Costa Rica, pp. 32-35.

[6] Cabral, A., & Goncalves, L. (2006). "SCORM Standard Compliant Learning Objects Generated from Content Organized with Concept Maps". in: A. J. Cañas & J. D. Novak (Eds.), *Concept Maps: Theory, Methodology, Technology. Proceedings of the Second International Conference on Concept Mapping*. San José, Costa Rica: Universidad de Costa Rica, pp. 519-526.

[7] Villalon, J. & Calvo, R. A. (2009), "Concept extraction from student essays, towards concept map mining", *in Proceedings of the Ninth IEEE International Conference on Advanced Learning Technologies*', ICALT '09, Washington, DC, USA, pp. 221–225.

[8] Lahti, L. (2009). Guided generation of pedagogical concept maps from the Wikipedia. Proc. World Conference on E-Learning in Corporate, Government, Healthcare and Higher Education (E-Learn 2009). 26-30 October 2009, Vancouver, Canada. AACE. ISBN 1-880094-76-2.

[9] Leake, D. B., Maguitman, A., Reichherzer, T., Cañas, A. J., Carvalho, M., Arguedas, M., & Eskridge, T. (2004, September). Googling from a concept map: Towards automatic concept-map-based query formation. In Concept maps: Theory, methodology, technology. Proceedings of the first international conference on concept mapping (Vol. 1, pp. 409-416).

[10] Wafula, B. N. A. (2006). Automatic Construction of Concept Maps. online:<<http://www.cs.joensuu.fi/pages/whamalai/sci wri/belinda.pdf.

[11] Karannagoda, E. L., Herath, H. M. T. C., Fernando, K. N. J., Karunarathne, M. W. I. D., de Silva, N. H. N. D., & Perera, A. S. (2013). Document Analysis Based Automatic Concept Map Generation for Enterprises. In *International Conference on Advances in ICT for Emerging Regions (ICTer)* (Vol. 154, p. 159).

[12] Lee, S., Park, Y., & Yoon, W. C. (2015). Burst analysis for automatic concept map creation with a single document. *Expert Systems with Applications*, 42(22), 8817-8829.

[13] Sun, Y. (2004). Methods for automated concept mapping between medical databases. *Journal of biomedical informatics*, 37(3), 162-178.s

[14] Luckie, D. B., Harrison, S. H., Wallace, J. L., & Ebert-May, D. (2008). Studying C-TOOLS: Automated grading for online concept maps. In *Conference Proceedings from Conceptual Assessment in Biology II* (Vol. 2, No. 1, pp. 1-13).

[15] Kumaran, V. S., & Sankar, A. (2013, October). An automated assessment of students' learning in e-learning using concept map and ontology mapping. In *International Conference on Web-Based Learning* (pp. 274-283). Springer, Berlin, Heidelberg.

[16] Nugumanova A., Mansurova M., Alimzhanov E., Zyryanov D., Apayev K. (2016) An Automatic Construction of Concept Maps Based on Statistical Text Mining. In: Helfert M., Holzinger A., Belo O., Francalanci C. (eds) Data Management Technologies and Applications. DATA 2015. Communications in Computer and Information Science, vol 584. Springer.

[17] Yevgeniy Bodyanskiy and Iryna Perova (2017), Medical Online Neuro-Fuzzy Diagnostics System with Active Learning, International Journal of Advances in Computer and Electronics Engineering, Volume 2, Issue 7, pp. 1–10

Author's Biography

Ramasamy Kumar, is a Professor of Department of IT in Rajalakshmi Engineering College, Chennai. He completed his Ph.D., in faculty of Information and Communication Engineering at Anna University, Chennai. His research interests include Cognitive Science, Data Mining, E-Learning and Knowledge Management.