Creating Knowledge Map for Topic Searching

Te Lin
School of Electronic Information and Electrical Engineering
Shanghai Jiao Tong University
Shanghai, China 200240
Email: 14310115150206@sjtu.edu.cn

***Abstract*—Knowledge is the primary asset of today's organizations; thus, knowledge management has been focused on discovery, representation, modification, transformation, and creation of knowledge within an enterprise. A knowledge map is a knowledge management tool that makes organizational processes more visible, feasible, and practicable. It is a graphical representation of decision-related information. What happens, how various events can be managed, and why they happened: all can be demonstrated very precisely by a well-designed knowledge map. When users are searching academic keywords of their interest in Acemap, they can obtain related search results, including papers, authors, affiliations, fields of study and other academic information. However, the information returned by the searching engineering is discrete, besides, users can't acquire whole field of the keywords they search, the relationship and hierarchy between different topics is unknown. Sometimes even when the users themselves don't know what topic the keywords are related to, it is necessary to give them a clear hierarchy map for the topics in the whole field. This paper proposes a new knowledge map for the representation of the keyword, including extracting fields of study, clustering the topics and visualization.**

I. Introduction

One of the common knowledge management and representation tools is a knowledge map, which is a visual representation of information. The concept originated from a metaphor for a geographical map. A knowledge map can be created at different levels of detail and allows experts to communicate with each other and share their knowledge. Generally, a knowledge map is considered a decision-support tool that can analyze knowledge. It provides context and identities an existing knowledge-base.

II. The Proposed Method

We propose three steps to achieve the knowledge map, extracting hierarchy information, control the scale of knowledge map and recursion.

A. Extracting Hierarchy Information

When users are searching academic keywords of their interest in Acemap, the searching engine solr will returned related papers and their topics. Our work is to ranking every topics frequency in the whole returning list. Once we acquire the topic list, for every topic, we search its father topics in the database, thus we can obtain the topic graph. The topic graph can reflect the relationship and hierarchy of different topics.

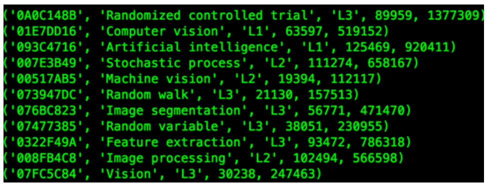

Fig. 1. Ranking the frequency of topics

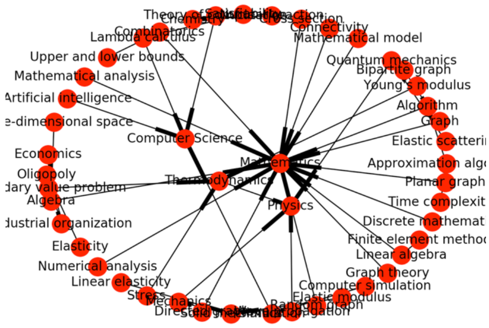

Fig. 2. Topic Graph

1) Ranking Topics: Ranking topics involves various factors including the frequency of fields of study, the citation, time, conference type, etc. What we should do is to integrate factors with different weight function and ranking the output of integration. Part of the topics' ranking is shown in Fig. 1.

2) Obtaining Hierarchy Through Database: After we obtain the topic list ranked in the previous step, for each topic we search its father topic, the father topic will be included as a new topic in the topic list. Then we search the newly included topic's father topic, we repeat this proceed until the topic has no father. Until now, we can obtain the topic graph which can demonstrate the relationship and hierarchy of different topics. The topic graph is shown in Fig. 2.

B. Control the Scale of Knowledge Map

Because of the immense scale of the topic graph, the knowledge map based on this topic graph will contain redundant information. We need to simplify the topic graph, which means controlling the scale of topic graph, extracting the primary information of it. The first step we should do is to change the form of topic graph G, we transform it to the topic tree T to

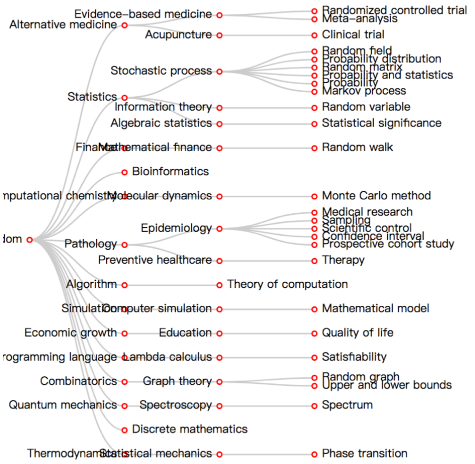

Fig. 3. Topic Tree

make the proceed clearer and more efficient. The transformed one is shown in Fig. 3.

1) Filter the Topic Preliminarily: For a topic tree T, if we delete the root node. We can obtain subtree sequences $\{T_1, T_2, \cdots T_M\}$. Root node in each subtree will be examined if the hierarchy of node is $'L_2'$ or $'L_3'$, we suppose every node having a hierarchy which represents its level of knowledge, the hierarchy set is $\{'L_0','L_1','L_2','L_3'\}$. If the hierarchy of node i is $'L_2'$ or $'L_3'$, the subtree T_i will be deleted. This method is to filter the topic which scale is two small and its hierarchy level can't represent the knowledge.

2) Clustering the Topic:

a) Computing the Scale: After we filtering the tree sequences $\{T_1, T_2, \cdots T_M\}$, we can now obtain a new filtered subtree sequences $\{T_1, T_2, \cdots T_N\}$, where $N \leq M$.
We define a scale function $S(T)$ of each T, according to the recursive property of tree, we can transform $S(T)$ into the recursive form below:

$$S^Q_{j-1,M(j-1)}(T) = \sum_{i=1}^{n} m^{M'(j)}_{j,i} \cdot S^{M'(j)}_{j,i}(T) + N_{j-1,M(j-1)} \tag{1}$$

where $S^Q_{j-1,M(j-1)}$ is the sum scale of node M in level $j-1$, node Q is the predecessor of node M. $S^{M'(j)}_{j,i}(T)$ represents the sum scale of all the descendent trees $\{M'\}$ of node M in level j. $m^{M'(j)}_{j,i}$ is the i_{th} transfer function in the descendent trees $\{M'\}$ in level j. $N_{j-1,M(j-1)}$ is the scale of node M in level $j-1$.
Now we give the expression of $m^{M'(j)}_{j,i}$, $N_{j-1,M(j-1)}$:

$$N_{j,M(j)} = \left(\alpha_1 L_{j,M(j)} + \alpha_2 \frac{1}{D_{j,M(j)}}\right) \cdot \frac{F_j}{\sum_{i=1}^{U} F_i} \tag{2}$$

where $L_{j,M(j)}$, $D_{j,M(j)}$ represent the length and depth information of node M in level j, α_1, α_2 is the weight coefficient of two information, where the expression of F_j is computed below:

$$F_j = \frac{f_j - f_{min}}{f_{max} - f_{min}} \tag{3}$$

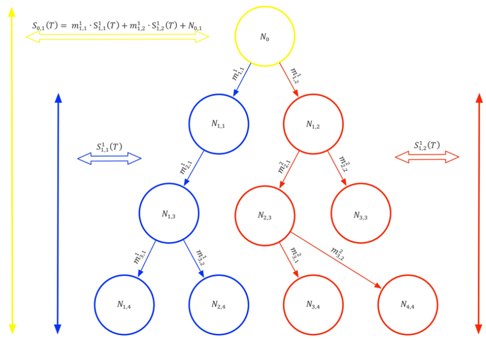

Fig. 4. Scale Computing

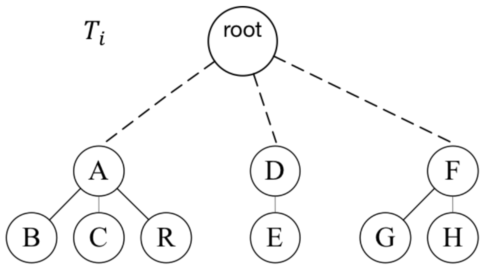

Fig. 5. Cutting Tree

where f_j is the j_{th} largest appearance frequency, f_{max} and f_{min} are the highest and lowest appearance frequency.
Finally, the expression of transfer function is shown below:

$$m^{M'(j)}_{j,i} = \frac{\delta}{\delta_{i,j}} \tag{4}$$

As illustrated in Fig. 4, the scale computing can be shown.

b) Clustering and Simplify Subtrees: Now we can obtain the scale sequences $\{S_1, S_2, \cdots S_N\}$, according to the sequences, we clustering the N-dimension vector with K-means method, and the parameter K is decided by us, the clustering result can be shown below:

$$\left\{S^1_1, S^1_2, \cdots S^1_{n_1}\right\}, \left\{S^2_1, S^2_2, \cdots S^2_{n_2}\right\}, \cdots \left\{S^k_1, S^k_1, \cdots S^k_{n_k}\right\} \tag{5}$$

We calculate the means of each cluster and choose the largest cluster, which means we choose $k = 1$, deriving the sequences $\left\{S'_1, S'_2, \cdots S'_{n'}\right\}$. Each of them correspond to the subtree sequences $\{T_1, T_2, \cdots T_\omega\}$, now we have selected the most important knowledge subtrees in the initial subtree sequences $\{T_1, T_2, \cdots T_M\}$.

3) Recursion: For the selected subtree sequences $\{T_1, T_2, \cdots T_\omega\}$, we delete each root of them, thus obtaining forest sequences, as illustrated in Fig. 5. For each forest, we can view the forest as a new set of tree sequences. We utilize the step **Clustering and Simplify Subtrees** and generalize the selected subtree sequences, repeating this proceed until the leaves.

C. Visualize the Knowledge Map

After we obtain the controlled topic graph, to make it more informative, we should integrate different variables

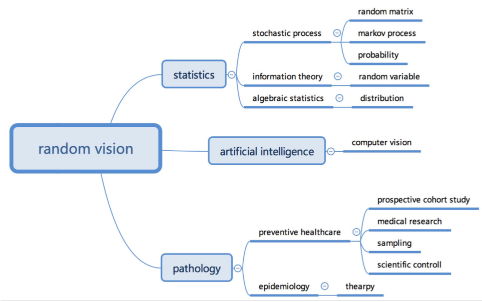

Fig. 6. Knowledge Map

including depth, size, color, fill of the stripe and frame when we are visualizing the topic graph.
The final knowledge map is shown in Fig. 6:

III. Future Work

The future work includes two parts:

a) Theoretical Part: In this paper, it uses only one dimension vector-scale function to describe the properties of tree T. However, there must have been other features to describe the properties of tree T. We are seeking useful features to be added into the feature vector to accomplish the clustering.

b) Visualization: The final knowledge map isn't informative now because the size of stripe and frame.

IV. Conclusion

The knowledge map illuminate both topic relation and hierarchy of what the user is searching, broaden and inspire user to think out of the box as well as offering a new method of interdisciplinary learning. Not only the predecessor and descendent topic is shown in the knowledge map, the clearer map of the whole field and the relationship between different fields are displayed.

V. Acknowledgement

I would like to appreciate all the fellows in AceMap group for building the website. Also my group leader Peng Qianyang has given me plenty of useful suggestion on data processing and theoretical part. At the visualization part, I got a lot of help from Zhao Jinhao, who showed me the framework of manipulating Gephi in Java. Their helps made it a lot easier for me to accomplish this project. And I will be specially thankful to Dr. Fu Luoyi who has instructed me on text clustering.