Visualization Tool for Finding of Researcher Relations

Takafumi Aoki[1], Yoshikazu Sasamoto[1], Keisuke Makita[2], and Shingo Otsuka[2]

[1] Kanagawa Institute of Technology, Atsugi-shi Kanagawa 243-0292, Japan
{s1121033,s1021076}@ccy.kanagawa-it.ac.jp
[2] Graduate School of Kanagawa Institute of Technology,
Atsugi-shi Kanagawa 243-0292, Japan
s1385022@cce.kanagawa-it.ac.jp, otsuka@ic.kanagawa-it.ac.jp

Abstract. It is possible to collect knowledge of interest research field effectively if we can look for the key person of the field. In addition, when the people who belong to administrations and companies want to undertake information gathering to the person of a particular field, it is convenient if the key person of the field is found easily. In this paper, we propose the visualization tool for finding of researcher relations using the conference programs.

1 Introduction

It is possible to collect the information about the field effectively if we understand the human relations in the field. Hence, we visualize the human relations in the specific field using conference programs, and we build the tool which can identify people with a strong relations. In general, conference programs are comprised of the session that put together the study of the same field. Each session is comprised of one chairperson and several presentations, and one or plural authors are listed in a presented paper. The papers are more likely to be collaborative research if the authors consist of several varying in affiliation. Therefore we can suppose that strong human relations are built. Moreover, there is some kind of relations between authors who presented in the same session because each session compiles the presentations of the same research field. And about chairperson, it is similar. We might extract human relations by using many conference programs. We extract the information of chairpersons and authors using the held program in the holding schedule of the IEICE and conference information that we obtained originally. Furthermore, we visualized human relations using the Graphviz[1] that AT&T Research provided and the JAVA.

2 Related Works

SPYSEE [2] is famous for a tool@checking human relations on webpages in Japan. Figure 1 shows an example of human relations in SPYSEE. This person is famous in Japanese companies. This person is investigated using information

S. Yamamoto (Ed.): HIMI 2014, Part I, LNCS 8521, pp. 3–9, 2014.

Fig. 1. Example of "SPYSEE"

of webpages and the relation graph is displayed by SPYSEE. However, this graph is wrong because there is Leonardo da Vinci to the friend of the friend of this person. There is doubt the authenticity of search results because this systems use information drawn from the webpages. On the other hand, there are many challenge of the extraction of researchers' relationships and the creation of the overhead view of research field using the bibliography information of the article [3,4,5,6,7,8]. In these studies, it extract the related information by using references and the author information of the articles. It is a subjective judgment because the references are recognized by authors to be the related studies. Our method is an objective judgment because we use the conference programs that the program chair compiles related study and divided into every session.

3 Acquisition of Research Group Data

We describe the authors list of that we acquired from the conference programs placed in the society pages. We also describe the Graphviz and the method of build process of "dot files".

3.1 Build Process of Authors' List Using Conference Programs

We extract the information of the session chair and the authors in the webpages of the conference programs for extracting researcher relations. We use two

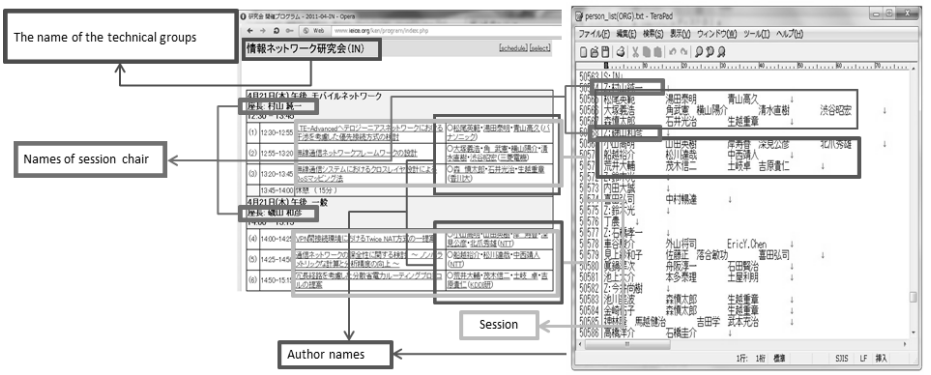

Fig. 2. Example of creating data set

dataset in our experiment. One is authors' list which we made hand-operated using the conference programs in conjunction with the database community in Japan, and the other is authors' list which we made one by manual operation using the conference programs of "the Institute of Electronics, Information and Communication Engineers (IEICE)"[9]. The website of IEICE has Technical Committee Submission System page[10]. Each information of conference schedule is made automatically as the left side of Figure 2. We extract authors and session chairs automatically using web scraping technology. We show a part of the list of authors on the right side of Figure 2. The list of persons in the right-side of Figure 2 shows following information.

- "S:" means a study group name.
- "Z:" means session chair(s).
- The line after "Z:" means authors. We express the multi-author using tab delimited.

3.2 Conversion from Authors' List to DOT Notation

We made a program to convert from the authors' list to DOT notation. The DOT notation is an expressing a graph with a text and describes it like Figure 3. It is easy to understand the human being while being the form that it is easy to computerize. We show the example which painted pictures using the Graphviz in Figure 4. We understand that "6" is connected from "1", "1" is connected from "4" and "2" is connected from "5". We use non-directed graph in our examination however it is possible to use directed graph in changing a parameter. And it is possible to change a color and size by appointing the attribute of the node and the edge.

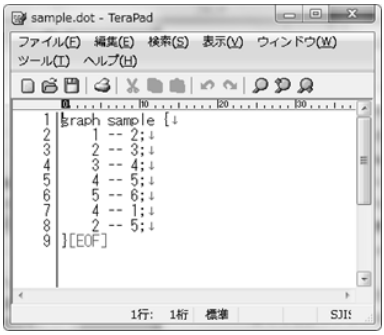

Fig. 3. Example of dot notation

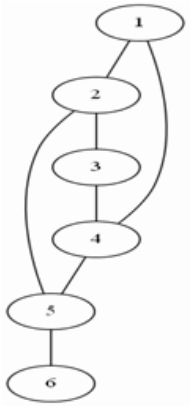

Fig. 4. Example of graphviz

3.3 Image Transformation Using Graphviz

The dot files convert picture files using the Graphviz. We express the associated person with "one to one" and express the strength of ties in "len" as follows in our experiments. The value of "len" grows big if the number of times of co-author, same session and relation with chairperson and presenter. It becomes the layout that is placed near if a value of "len" is big. And we expressed the strong relation in a red line because there are many displayed number of nodes (researchers).

graph "g" "Takafumi Aoki" – "Keisuke Makita" [len = 42.00]; "Keisuke Makita" – "Yoshikazu Sasamoto" [len = 41.00];

4 Experimental Result

4.1 Details of Dataset

We use the conference programs written in Japanese in ours experiments. Dataset A is conference program data of the workshop and the forum for 41 times related to the databases community in Japan. Dataset B added data set A to the conference program data for 2,306 times which we are able to acquire from 2,457 links in "the technical Committee Submission System pages" in the IEICE website. And we succeed in collecting the 60,000 researchers' names. We visualize human relations of two dataset using the Graphviz and the graph tool which we made by JAVA.

4.2 Visualization Examples

We show the example of the Graphviz visualization using dataset A in the Figure 5. We confirm that a researcher tied to a great many people in the center

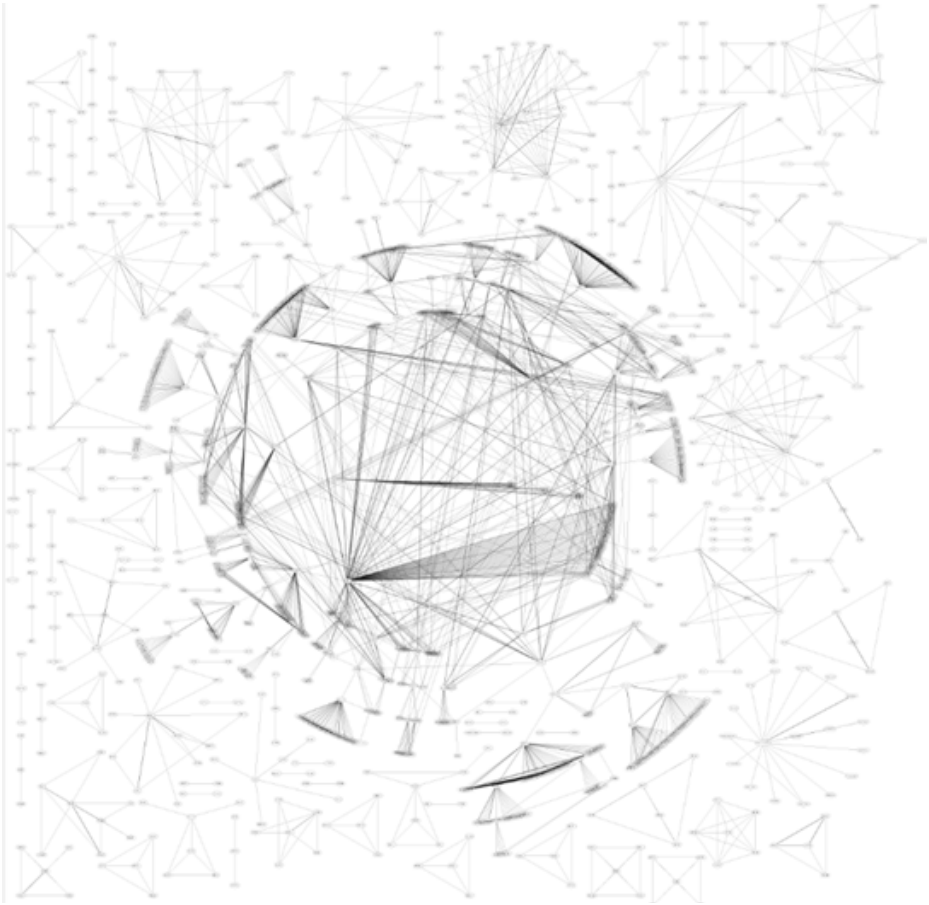

Fig. 5. Result of visualization(1)

of Figure 5. In fact, this researcher is one of the key person in the database community and the many pupils of this researcher play an active part. There are many sectors in a Figure 5. The sectors express master and pupil relations and a sector tends to be big as laboratory of the large family.

Next, we show the example of our original tool using dataset A in the Figure 6. There is a slide bar changing the number of the nodes (researchers) under the center in the Figure 6. The edges are displayed if the degree of relation between nodes (researchers) is high. And the edges are short as the degree between nodes relation is high. The position of each node does not have the meaning. Each node is placed at random first. When time passes, the nodes that are high in a degree of relation become near. It is possible to fix the node if the users click the left button of a certain node and the color of the nodes related to the node changes if the users click the right button. The example in the Figure 6, we understand

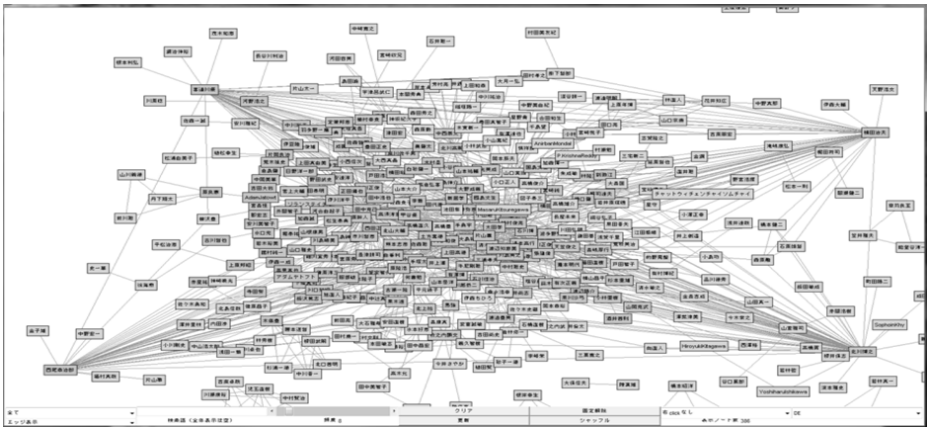

Fig. 6. Result of visualization(2)

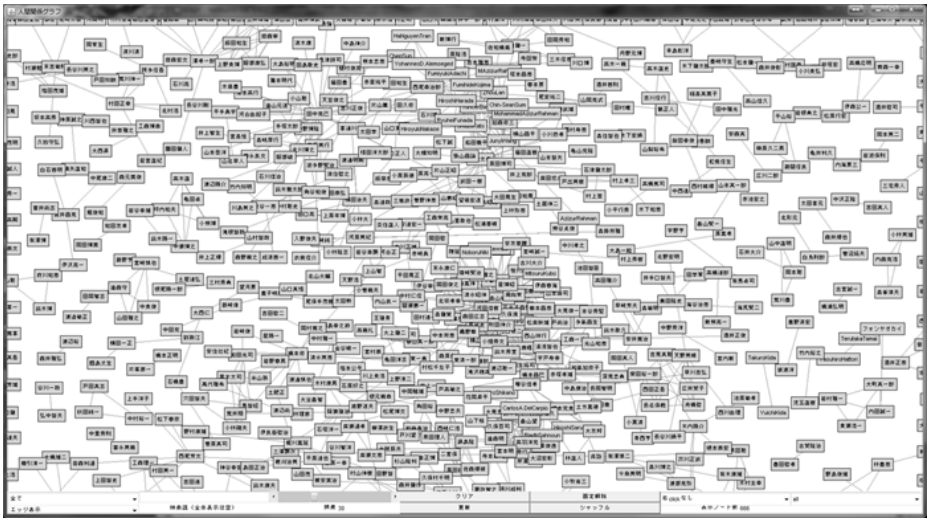

Fig. 7. Result of visualization(3)

that the nodes are many and gathered in a circular pattern as this node is the well-known researcher of the database community.

Finally, we show the example which we visualize with our tool using dataset B in Figure 7. We understand that it becomes the form that scattered in whole although there is the place where the nodes interval becomes dense. The dataset B includes approximately 60,000 researchers. With 27 inches of display which we used by the experiments, indication of about 1,000 nodes is a limit. So we make frequency a considerably high value. Therefore, the most of the nodes displayed here seem to be the well-known researchers.

Figure 5 and Figure 6 shows an example about a particular technical group. The color of the nodes in conjunction with the selected node changes. It is possible

to understand the relation between researchers in detail. However, the number of nodes that the user with our tool can grasp depends on the screen size.

4.3 Discussion

The relations of the researchers are captured for overlooking using the Graphviz in experimental results. And our tool can grasp the relations between researchers in greater detail. As problems of the present, we must thin out the number of the nodes to display beforehand because the node number that the Graphviz can express has a limit. Moreover, the correspondence to the dynamic demand of the users is difficult because the Graphviz makes a conversion to an image. In our original tool, the improvement of the point that the number of nodes that the users can grasp depends on display size for is necessary.

5 Conclusion

We try to visualize human relations using the conference programs. We also extract researchers' information using the society webpages and visualize human relations about a particular field. We will perform the visualization and the comparison in other research fields in future.

References

1. Graphviz, `http://www.graphviz.org/Home.php`
2. SPYSEE, `http://spysee.jp/`
3. Kondo, T., Nanba, H., Takezawa, T., Okumura, M.: Technical trend analysis by analyzing research papers' titles. In: Vetulani, Z. (ed.) LTC 2009. LNCS, vol. 6562, pp. 512–521. Springer, Heidelberg (2011)
4. Nanba, H., Kondo, T., Takezawa, T.: Automatic creation of a technical trend map from research papers and patents. In: Proc. the 3rd International CIKM Workshop on Patent Information Retrieval (PaIR 2010), pp. 11–15 (2010)
5. Fukuda, S., Nanba, H., Takezawa, T.: Extraction and visualization of technical trend information from research papers and patents. In: Proc. the 1st International Workshop on Mining. Scientific Publications (2012)
6. Karamon, J., Matsuo, Y., Yamamoto, H., Ishizuka, M.: Generating social network features for link-based classification. In: Kok, J.N., Koronacki, J., Lopez de Mantaras, R., Matwin, S., Mladenič, D., Skowron, A. (eds.) PKDD 2007. LNCS (LNAI), vol. 4702, pp. 127–139. Springer, Heidelberg (2007)
7. Nguyen, M., Kato, D., Hashimoto, T., Yokota, H.: Research history generation from metainformation of research papers using maximum margin clustering. International Journal of Business Intelligence and Data Mining 7, 217–231 (2012)
8. Jin, Y., Matsuo, Y., Ishizuka, M.: Extracting social networks among various entities on the web. In: Franconi, E., Kifer, M., May, W. (eds.) ESWC 2007. LNCS, vol. 4519, pp. 251–266. Springer, Heidelberg (2007)
9. The Institute of Electronics, Information and Communication Engineers, `http://www.ieice.org/eng/index.html`
10. IEICE Technical Committee Submission System, `http://www.ieice.org/ken/program/index.php?lang=eng`