

Journal Visualization by a Dual Hierarchical Data Visualization Technique

Kana Shiratori*
Ochanomizu University

Takayuki Itoh†
Ochanomizu University

1 INTRODUCTION

Researchers often search for specific papers from journals, or look over the trends of the journals. This task is not always easy: for example, various kinds of papers may be extracted if we use general terms for query operations, and it may prevent to discover the interesting papers. We think journal visualization is useful to look the distribution of queried papers or overview of the journal itself. This poster presents a journal visualization technique, applying our dual hierarchical data visualization technique [2]. The technique firstly extracts meaningful keywords from a journal, and calculates importance of the keywords for each paper or author. It finally represents the distribution of keywords, papers, and authors by our own dual hierarchical data visualization technique. The poster also provides our experiments that demonstrate the effectiveness of the presented technique.

2 RELATED WORK

Many of document visualization techniques represent clusters or relationships of documents constructed based on the co-occurrence of keywords [1]. Our technique in this poster is also based on the co-occurrence of keywords. DualNAVI [4] is one of the existing techniques most related to our work, which represents relationship of keywords in the left side of a window, and a list of documents in the right side of the window.

We proposed a hierarchical data visualization technique [3], which represents leaf-nodes of hierarchical data as colored icons, and branch-nodes as nested rectangular regions. Our dual visualization technique [2] applies this hierarchical data visualization from side to side, to effectively represent matrix data. The technique firstly applies a clustering algorithm to columns and rows of matrix data independently, and generates two hierarchical data as the clustering results. The technique then displays the two hierarchical data respectively, while it provides a user interface so that users can interactively explore the two hierarchical data each other. We have applied this technique to visualization of newspaper corpus data that represents the relationship among keywords and articles of the newspaper. In addition to the representation of keywords and documents, the technique presented in this poster also focuses on representation of authors.

3 JOURNAL VISUALIZATION

Figures 1 and 2 show snapshots of journal visualization by the presented technique. The technique visualizes hierarchically clustered authors and papers by our dual hierarchical data visualization technique, and displays keyword buttons in the left end of the display space.

3.1 Extension of our dual hierarchical data visualization

Our previous work [2] visualized newspaper corpus data, by constructing a 2D keyword-article matrix and filling the importance

*e-mail: kana@itolab.is.ocha.ac.jp

†e-mail: itot@is.ocha.ac.jp

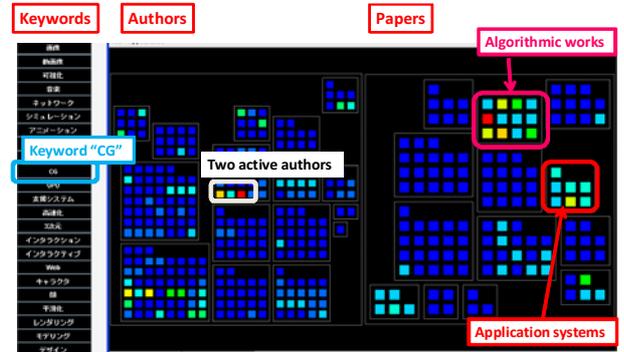


Figure 1: Snapshot of the presented technique (1). A keyword "CG" is selected.

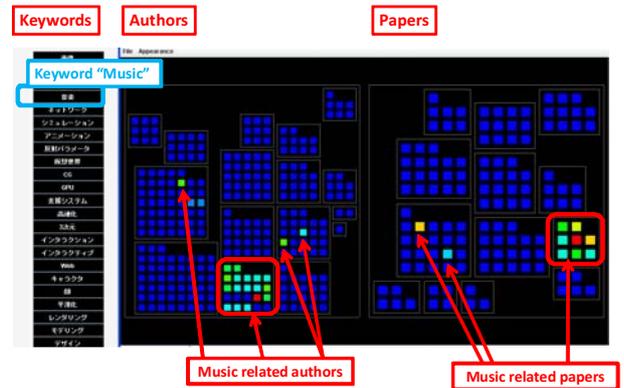


Figure 2: Snapshot of the presented technique (2). A keyword "Music" is selected.

values in. The technique presented in this poster extends the mechanism by adding author information. It constructs a 3D keyword-paper-author matrix filling the importance values in, and then visualizes the 3D matrix.

3.2 Processing flow of journal visualization

Figure 3 shows the data structure and processing flow we suppose in this poster. Here, we describe the keyword-paper matrix as T_1 , and the keyword-author matrix as T_2 . We also describe papers as r_1 to r_n (n is the total number of papers), authors as s_1 to s_l (l is the total number of authors), and keywords as c_1 to c_m (m is the total number of keywords). In the matrix data, a_{ij} denotes an importance of the j -th keyword in the i -th paper, and b_{kj} denotes an importance of the j -th keyword of the k -th author.

The technique firstly applies a clustering algorithm to papers and authors independently. It then displays hierarchical clusters of papers in the right side of the window, and hierarchical clusters of authors in the left side of the window. It also displays the keywords as a vertically aligned set of buttons. Colors of icons and buttons denote the importance values. The importance is high if the color

is close to red. On the other hand, it is low if the color is close to blue.

3.3 Interaction among keywords, authors, and papers

The presented technique has a mechanism to click keywords, papers, and authors, so that users can interactively explore the matrix data. For example, when a user clicks a keyword button, icons of papers and authors related to the clicked keyword are highlighted. Similarly, when a user clicks an icon of the paper/author, icons of authors/papers and keyword button related to the clicked paper/author are highlighted.

Keyword operation

Suppose that a user clicks a keyword button c_j . The technique scans the values a_{1j} to a_{nj} in the matrix T_1 , and uses the scanned values a_{ij} to calculate colors of icons of papers r_i . Similarly, the technique scans the values b_{1j} to b_{lj} in the matrix T_2 , and uses the scanned values b_{kj} to calculate colors of icons of authors s_k .

Reflection of author/paper operation to keywords

Suppose that a user clicks an icon of a paper r_i . The technique scans the values a_{i1} to a_{im} in the matrix T_1 , and uses the scanned values a_{ij} to calculate colors of keyword buttons c_j . Red lines and circles in Figure 3 denote this processing flow. The technique applies the similar mechanism when a user clicks an icon of an author s_k .

Reflection of author/paper operation to papers/authors

Suppose that a user clicks an icon of a paper r_i . The technique calculates the color of the icon of the k -th author s_k from the value $\sum_{j=1}^m a_{ij}b_{kj}$, which corresponds to the sum of importance of all keywords of the k -th author b_{kj} weighted by the importance of the clicked i -th paper a_{ij} . Blue lines and circles in Figure 3 denote this processing flow. The technique applies the similar mechanism to calculate colors of icons of papers, when a user clicks an icon of an author s_k .

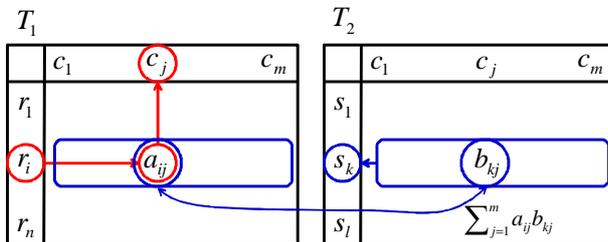


Figure 3: Processing flow of color calculation with click operations.

4 EXPERIMENT

We applied a new journal "Journal of the art and science"¹ published in Japan since 2002 as test data of the presented technique. In this experiment, we firstly extracted abstract and author information from the all 134 papers of the journal. At this moment 283 authors have contributed to the journal. We then selected important keywords from the abstracts, by using Japanese language morphological analysis software "Chasen"², and word importance calculation software "termex"³. In our experiment we automatically selected top 40 keywords, and then manually selected especially meaningful 23 keywords from the 40 keywords.

¹<http://art-science.org/journal/index.html>

²<http://chasen.naist.jp/hiki/Chasen/>

³<http://genshen.dl.itc.u-tokyo.ac.jp/>

Figure 1 shows a visualization result when a user selects a keyword "CG". The left part of the visualization result displays especially bright (red and orange) two icons in a central cluster. It denotes that the two authors especially contributed to the journal by submitting CG-related papers. The right part of the visualization result shows that there are two clusters (marked by pink and red circles) that contain especially bright icons. The upper cluster marked by pink contains papers highly related to other several keywords, "modeling", "3D", and "image". The lower cluster marked by red contains papers highly related to other several keywords, "support system" and "design". The result denotes that the presented technique successfully divided CG-related papers into two clusters of algorithmic papers and application system papers, and the visualization technique successfully represented the two clusters. We think that such representation helps users to focus on interested clusters of the papers.

Figure 2 shows a visualization result when a user selects a keyword "Music". The left part of the visualization result shows that most of music-related authors construct a single cluster. On the other hand, several other music-related authors are out of the cluster; they are in the clusters which consist of authors of image- or visualization-related papers. The right part of the visualization result also shows that most of music-related papers construct a single cluster; however, several other papers are out of the cluster. Actually, most of music-related papers in the journal are related each other, while the papers present music technologies such as composition assistance system and sound editing algorithm. On the other hand, several papers are also related to other keywords; actually one of the papers presents an algorithm for matching between images and tunes. The visualization result denotes that users can discover outlier papers related to specific keywords ("Music" in this case) by the presented technique.

5 CONCLUSION

The poster presented a journal visualization technique applying our dual hierarchical data visualization technique. We would like to have experiments with larger journal data as a future work.

Also, we would like to extend the functionality of the technique. For example, we would like to represent visibility of authors and papers as 3D bars so that users can easily focus on popular authors or papers.

This work has been partially supported by Japan Society of the Promotion of Science under Grant-in-Aid for Scientific Research (C) No. 18500074.

REFERENCES

- [1] K. Andrews, W. Kienreich, V. Sabol, J. Becker, G. Droschl, F. Kappe, M. Granitzer, P. Auer, K. Tochtermann, The InfoSky Visual Explorer: Exploiting Hierarchical Structure and Document Similarities, *Information Visualization*, Vol. 1, No. 3, pp. 166-181, 2002.
- [2] T. Itoh, H. Tachibana, Visualization of Corpus Data by a Dual Hierarchical Data Visualization Technique, *The 7th NTCIR (NII Test Collection for Information Retrieval Systems) Workshop Meeting on Evaluation of Information Access Technologies*, 534-541, 2008.
- [3] T. Itoh, H. Takakura, A. Sawada, K. Koyamada, Hierarchical Visualization of Network Intrusion Detection Data in the IP Address Space, *IEEE Computer Graphics and Applications*, 26(2), 40-47, 2006.
- [4] A. Takano, Y. Niwa, S. Nishioka, M. Iwayama, T. Hisamitsu, O. Imaichi, H. Sakurai, Associative Information Access Using DualNAVI, *Kyoto International Conference on Digital Libraries (ICDL'00)*, pp.285-289, 2000.