Muse: Visualizing the origins and connections of institutions based on co-authorship of publications

Till Nagel[1,2] , Erik Duval[1]

[1] Dept. Computerwetenschappen, Katholieke Universiteit Leuven, Celestijnenlaan 200A, 3001 Heverlee, Belgium
[2] Interaction Design Lab, University of Applied Sciences Potsdam, Pappelallee 8-9, 14469 Potsdam, Germany
nagel@fh-potsdam.de, erik.duval@cs.kuleuven.be

Abstract. This paper introduces Muse, an interactive visualization of publications to explore the collaborations between institutions. For this, the data on co-authorship is utilized, as these signify an existing level of collaboration. The affiliations of authors are geo-located, resulting in relations not only among institutions, but also between regions and countries. We explain our ideas behind the visualization and the interactions, and briefly describe the data processing and the implementation of the working prototype. The prototype focuses on a visualization for large tabletop displays, enabling multiple users to explore their personal networks, as well as emerging patterns in shared networks within a collaborative public setting. For the prototype we used the publication data of the EC-TEL conference.

Keywords: geo-visualization, tabletop, research, human computer interaction

1 Introduction

There has been vast amount of research in the areas of bibliometry and scientrometry to extract and specify the metrics of scientific publication and citation networks. Several works used approaches to visualize these networks (e.g. [1], [2]). In the field of TEL, [3] analyzed and visualized ED-Media publications.

The objective of the presented visualization is not to study individuals and their personal co-authorship networks, but rather to enable analyzing the connection network of universities and research centers. The inter-institutional relationships are based on co-author data, as "co-authorship seems to reflect research collaboration between institutions, regions, and countries in an adequate manner" [4].

Our intention is to focus attention on the spatial relations by creating an easy-to-understand geo-visualization with an emphasis on affiliations and collaborations between these institutions. Studies have shown geographic proximity is important and does positively influence the intensity and frequency of scientific collaboration [5]. However, there has been little research on using geo-visualization for inter-institutional and inter-country collaboration based on publication data (e.g. [6]).

This work focuses on an interactive geo-visualization on large display, enabling multiple users to explore the networks of their affiliations, as well as emerging patterns in shared networks within a collaborative public setting.

We envision several use cases for the application, from which we briefly describe three, exemplarily. (1) A visitor wants to get an overview of the spatial characteristics of scientific collaboration. He starts exploring the institutions and their locations, with the application showing the number of co-authored publications over the years. This visualization supports him understanding whether there is a correlation between proximity and the amount of collaboration. (2) An attendee is interested in finding future partners for writing a proposal. She sees that a colleague from her institution once co-authored a paper with someone from a university department in her field. She writes down the author's name, to later ask her colleague to introduce her. (3) Two persons stand at the table and both are exploring their own affiliations. The application highlights the respective publications, thus enabling them to see shared publications of colleagues, by serendipity. They start talking about these former projects, and find out they have mutual research interests.

The paper introduces *Muse*[1], a working prototype, whose main purpose is to ease the exploration of collaborations between institutions. In addition, the use of a large display tabletop, as well as the aimed-for simplicity of visualization and interaction intend to invite attendees to participate, and engage in discussions at a conference location. The following chapter gives a short overview on the data set. A description of the prototypes' visualizations and interactions follows. The paper closes with short conclusions and comments on future work.

2 Data Set

We are using the EC-TEL dataset as first illustration to show the connectivity in the scientific TEL community. With a young conference as EC-TEL we will not be able to show long-term transformations. Instead, here our aim is twofold: Showing how a striving conference evolved over recent years, and enabling attendees to explore their scientific neighborhoods in the TEL domain.

We harvested the publication data from the website of Springer, the proceedings publisher. We used Web-Harvest [7] to collect all titles, authors, and affiliations including their postal addresses (as well as further data). As the data originally is provided by the authors, using various languages, formats, and accuracies of data, we needed to apply different aggregation and unification heuristics, trying to reduce unintentional duplicates or other skewed data entries. First, the affiliation line is split up into the affiliation's name and its address, to allow a better unification of affiliations, and to display a shorter and more readable name in the visualization. The simplistic, language agnostic approach was to concatenate all text segments up to and including the last segment containing one of a set of specific keywords, selected for high probability of matching institutional name segments (e.g. "universi",

[1] The name of the application was chosen to reflect the meaning of "to look thoughtfully at". Secondarily, Muse, the greek goddess, presides over literature and science.

"a[clk]adem"). Second, the affiliations were to be unified based on the similarity of the name[2]. After geo-coding the addresses, we also incorporated the spatial proximity to ensure not unifying institutions with very similar names but different locations, e.g. "Dept. of Preventive Medicine, Korea University, South Korea" and "Dept. of Preventive Medicine, Konkuk University, South Korea".

Generally, it is difficult to structure real-world objects in a way to map all possibilities and special cases, thus we utilized a good-enough approach. Before realizing the prototype we probed into the data and looked for patterns to establish the visualization will be able to reflect those inherent relationships. Some of our analysis for the EC-TEL conferences 2006-2009 can be found at [8].

3 Prototype

We designed two working prototypes, with an iterative development approach to refine the visualizations, and to increase the usability of the interactions. The first interactive visualization was presented at the Science2.0 for TEL workshop at EC-TEL 2009. The presentation, and the public display at the venue thereafter allowed us to gather informal responses of attendees. We tried incorporating the given feedback into the second version, and aimed for improving the clarity of the visualization and the overall user experience in an on-location conference setting.

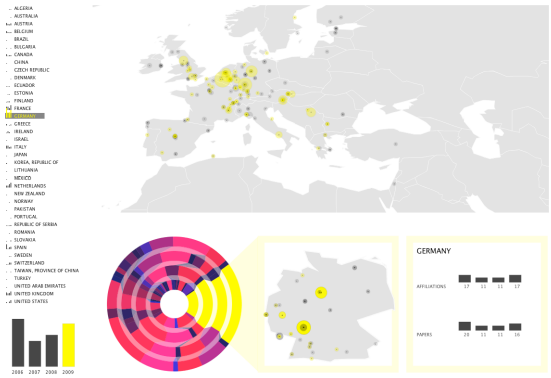

Fig. 1. Screenshot of first prototype with Germany and 2009 as selected country and year.

The first application consists of a static world map showing institutions as colored circles with its overall publication number mapped as size (see Fig. 1). Several further visualizations are in juxtaposition: An overview list shows the names of the countries with contributing authors, with a small *sparkline* [9] signifying the absolute

[2] This simplistic approach results in some false positives (e.g. "Av. Universidad 30"), which are recognized as part of the name, and some false negatives (e.g. "ETH Zürich"), which are regarded as part of the address. Furthermore, some entries could not be unified automatically, such as "Lehrstuhl Informatik V" with "Informatik 5 (Information Systems)".

publications over the years. The concentric rings represent the relative distribution of publications of every participating country over the years, starting from older (inner) up to the latest conference (outer).

These multiple displays are connected, and every user interaction is reflected in all other views. After selecting a country in one of the displays the application provides details-on-demand on that country, and its respective publications and institutions in simple bar diagrams. When the user selects a year the publications are filtered to highlight the data of that specific conference (i.e. as yellow circles and bars).

While the multiple displays allowed looking into the dataset from different perspectives, they also tended to clutter the screen. To effectively communicate the data in a concise visual manner some of the useful but distracting displays have been eliminated in the second prototype. The main improvements were to reduce the visual and interaction complexity by focusing on one main visualization, and the employment of an interactive tabletop with the aim to facilitate multi-user scenarios.

With the large interactive surface, the user not only views and manipulates data on a single user system, but operates in a collaboratively created and used information space (see Fig. 2). In this setting, co-located users, who may or may not be associated with each other, explore the visualization together. Users can arrive or leave at any time, and have the ability to interact as an individual, or as a member of a group with similar interests, goals or attitudes. Cooperative interaction can involve periods of tightly coupled activities by groups with similar but diverging goals, alternated with more loosely coupled individual work. Such collaborative threads can close, split off and merge repeatedly.

Fig. 2. Users exploring institutions with the tabletop prototype.

A single large world map showing all institutions and their relations based on co-authorship are displayed. The user is able to select the region she is interested in by panning and zooming the map (while in the first prototype a user only could switch between World and Europe). Even though more complex map manipulations are possible, we chose this interaction approach, as by reducing the prototype to a single visualization the user can concentrate on the map, thus lessening her efforts. The user can select a country she is interested in. That country is selected, and additional

information and diagrams are shown, similar as in the first prototype. These info-windows can be moved to any point on the table. When two countries are selected the prototype displays the diagrams besides each other, allowing the user to compare them.

4 Conclusion and Future Work

Although we have utilized only a small dataset with a rather small significance for general scientific network analysis, we see the Muse prototype with the used data set as beneficial case study. Through interactive filtering the user is able to explore the temporal as well as spatial relations between institutions, and can gather insights into the conference. The collaborative usage of the interactive tabletop display fosters communication among participants.

We intend to broaden the data set to other conferences. Currently, we see two possibilities: Besides using the harvesting tool to scrape further publications from Springer and other official sources, we plan to integrate publication data services, such as pub.fm [10]. Second, querying Web2.0 applications such as Mendeley [11] to gather social network data of the authors.

Furthermore, we are planning an evaluation on intelligibility of the visualization, and usability of the interactions. As direct response from users in a real-world setting can be worthwhile, we intend to create a brief questionnaire to gather feedback from attendees at the EC-TEL 2010.

References

1. Ponds, R., Van Oort, F. G., and Frenken, K. The geographical and institutional proximity of research collaboration. Papers in Regional Science, 86(3), 423–443. (2007)
2. Henry, N., Goodell, H., Elmqvist, N., Fekete, J.–D.: 20 Years of Four HCI Conferences: A Visual Exploration. International Journal of Human-Computer Interaction, 23:3, 239-285. (2007)
3. Ochoa, X., Méndez, G., Duval, E. Who We Are: Analysis of 10 Years of the ED-MEDIA Conference. In: G. Siemens and C. Fulford (eds.), Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2009, 189-200. Chesapeake, VA: AACE. (2009)
4. Glänzel, W., Schubert, A.: Analyzing Scientific Networks through Co-authorship. In: H.F. Moed et al. (eds.), Handbook of Quantitative Science and Technology Research, 257-276. Kluwer, Dordrecht (2004)
5. Katz, J. S. Geographical Proximity and Scientific Collaboration. Scientometrics 31 (1), 31–43. (1994)
6. Cyberinfrastructure for Network Science Center. Research Collaborations by the Chinese Academy of Sciences. http://scimaps.org/maps/map/research_collaborati_110/ (2009)
7. Web-Harvest Project. http://web-harvest.sourceforge.net/
8. Nagel, T. http://tillnagel.com/2010/09/ectel/
9. Tufte, E. Beautiful Evidence. Graphics Press. (2006)
10. Stellar project. http://www.stellarnet.eu/d/6/3/KULDocumentation
11. Mendeley. http://www.mendeley.com/oapi/