Visualization According To Research Paper Keywords

Petra Isenberg* Inria
Tobias Isenberg† Inria
Michael Sedlmair‡ University of Vienna
Jian Chen§ University of Maryland Baltimore County
Torsten Möller¶ University of Vienna

ABSTRACT

We analyzed visualization paper keywords supplied for 4366 papers accepted to three main visualization conferences. We describe main keywords, topic areas, and 10-year historic trends from author-chosen keywords for papers published in the IEEE Visualization conference series (now called IEEE VIS) since 2004. Furthermore, we present the KeyVis Web application that allows visualization researchers to easily browse the 2600+ keywords used for IEEE VIS papers of the past 10 years, aiming at more informed and, hence, more effective keyword selections for future visualization publications and efficient search for related work.

1 INTRODUCTION

The field of visualization is an extremely diverse and, thus, exciting field to be a part of. Yet, its diversity also creates enormous challenges. There are different levels of appreciation for all aspects of visualization research, communication challenges between visualization researchers, and the challenge of communicating visualization as a research science to the outside. These issues lead, in particular, to the frequently asked question "what is visualization?"—among funding agencies or even between colleagues. Yet, while "what is visualization?" is a fundamental question, it is little discussed within our community. What is particularly missing in this picture is a bottom-up analysis: What types of visualization research are actually happening as expressed by single research contributions in the visualization conferences and journals. Our analysis is one of the first steps in this direction. We analyze author-assigned keywords from the three IEEE VisWeek/VIS conferences of the past ten years and based on this analysis, we contribute:

- a conceptual map of all visualization work as indexed by individual authors
- KeyVis, a Web-based search tool that makes the keyword metadata available to a broad set of people.

2 RELATED WORK

Similar to our goal, in other disciplines specific techniques have been used to analyze the scientific literature more broadly: to get a better sense of global research trends, links and patterns within the scientific literature. Co-word analysis is one approach among others (e. g., co-citation analysis) that has tackled the problem by analyzing the scientific literature according to the co-occurrence of keywords, words in titles, abstracts, or even in the full texts of scientific articles [2]. Co-word analysis has been used in different research areas, e. g., polymer chemistry [1], acid rain research [3], or education [5]. Liu et al.'s work [4] is most related to ours. The authors examined papers of the ACM CHI conference from 1994–2013, identified research themes and their evolution, and classified individual keywords as popular, core, or backbone. We employ similar approaches but extend the work by an expert coding process. We also naturally differ as our focus is on a different research community with different keywords, trends, and patterns and a different historical evolution.

*e-mail: petra.isenberg@inria.fr
†e-mail: tobias.isenberg@inria.fr
‡e-mail:michael.sedlmair@univie.ac.at
§e-mail:jichen@umbc.edu
¶e-mail:torsten.moeller@univie.ac.at

3 CO-WORD ANALYSIS METHODOLOGY

For our analysis of the visualization research literature we collected keywords freely assigned by the authors to their research papesr accepted to IEEE VisWeek/VIS from 2004–2013. The dataset contained 1097 published papers (excluding posters). Out of these, 58 contained no author-assigned keywords, yielding a set of 1039 papers we considered in our analysis. These papers contained a total of 2823 unique keywords. Next, we engaged in an extensive manual cleaning pass in which we consolidated keywords that were the same but presented either as singulars/plurals, with spelling mistakes, or as acronyms. This yielded a cleaned dataset that contained 2629 unique keywords. Based on this data, we engaged in a manual expert coding in order to find higher-level clusters of keyword topics. The resulting set of keywords contained 156 unique higher-level keywords that occurred a total of 4026 times across all papers (this number is lower than for the cleaned data as potentially generated duplicates per paper were removed).

To analyze the keyword datasets we first filtered the data and and removed keywords that occurred less than ten times and also excluded higher-level terms (visualization, information visualization, scientific visualization, visual analytics). Next, we generated document-keyword matrices and performed a correlation computation. On each resulting correlation matrix we performed a hierarchical clustering using Ward's method and a squared Euclidean distance metric. We also generated a keyword network in which two keywords were linked if their correlation was > 0 and each link was assigned its respective correlation value.

4 RESULTS

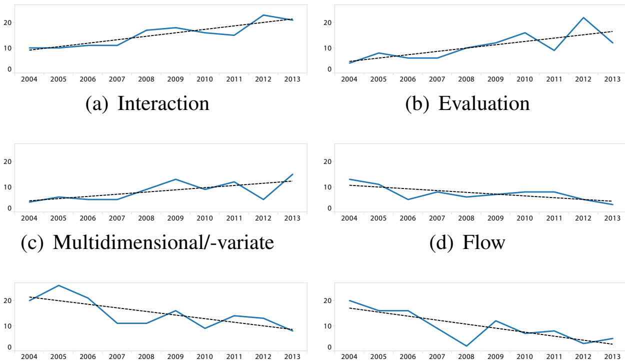
4.1 Individual Keywords

The top three most frequently occurring keywords across all three conferences were volume rendering (148×), hardware and computation (89×), and graphs (73×). Interestingly, there was little overlap between the most common keywords per individual conference. *Interaction* emerged for both IEEE InfoVis and VAST as the only shared keyword amongst the top three per conference:

VAST: *analysis process* (53), *interaction* (31), *applications* (24)
InfoVis: *graphs* (73), *interaction* (69), *evaluation* (53)
Vis/SciVis: *volume rendering* (148), *hardware and computation* (89), *flow* (70)

When expanding the analysis to the top 10 keywords, *application* emerges as a keyword common to all three conferences. InfoVis and VAST also share the keyword *evaluation* (ranked 14 in Vis/SciVis) while VAST and Vis/SciVis share the keyword *time* (ranked 12 in InfoVis). We can conclude that all three sub-communities share some joint concerns (e. g., *applications*, *evaluation*, *time*) but that they do have their respective foci.

Next, we were interested in historic trends for individual keywords. We used Tableau to calculate linear trend lines for the 15



(a) Interaction (b) Evaluation

(c) Multidimensional/-variate (d) Flow

(e) Volume Rendering (f) Hardware & Computation

Figure 1: Keywords significantly on the rise or in decline.

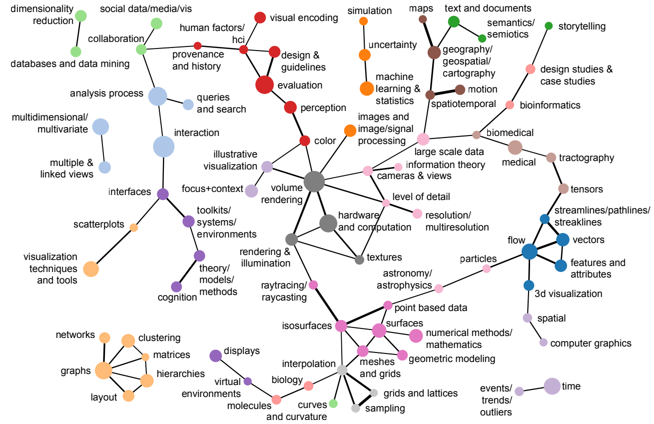


Figure 2: Keyword map from clustering of the coded keywords; showing only connected nodes with correlation strength ≥ 0.11. Circle areas correspond to number of occurrences of the keyword in the dataset, link width corresponds to correlation strength, and color distinguished node clusters. We chose a link threshold based on visual inspection of the resulting graph to generate a manageable and readable layout. Isolated nodes were removed from the image.

most frequently used keywords. According to this calculation, the terms *interaction*, *evaluation*, *multidimensional/multivariate*, and *machine learning & statistics* are significantly on the rise. In contrast, *volume visualization*, *flow visualization*, and *hardware and computation* have been significantly decreasing in frequency of occurrence over the past 10 years (see Fig. 1). It is interesting to note that two very core and frequent keywords for the IEEE Vis/SciVis conference are in significant decline. This could perhaps be due to the fact the many fundamental research questions have been tackled and that researchers are now using more specific or other keywords.

4.2 Keyword Clusters

Next we were interested to see which keywords were often used together to index papers and would, thus, describe the relatedness between visualization topics. We used the methodology outlined in Sect. 3 to derive topic clusters and network diagrams. Fig. 2 shows the keyword map in which keywords are connected if their correlation was ≥ 0.11, meaning that keywords were frequently used together to index individual papers. Node colors further indicate cluster memberships. In the following, we describe all clusters by their two most common keywords and the color used in Fig. 2.

Based on analyzing clusters with high density (keywords closely connected to one another) and centrality (many connections to other clusters) within the network [2], we found that three motor themes emerged for the community based on high density and centrality: 1–dark pink: surfaces, numerical methods/mathematics, 2–dark blue: flow, topology, 3–gray: volume rendering, hardware and computation. All three themes are traditional topics of IEEE Vis/SciVis which also has the longest history of the three IEEE VIS conferences.

Undeveloped but central themes in the network were characterized by high centrality and low density: 1–purple: displays, toolkits/systems/environments, 2–light green: data and data management, dimensionality reduction, 3–light pink: large scale data, cameras & views, 4–light purple: time, focus+context.

5 KEYVIS.ORG: A Keyword Search Tool

To make our data accessible for others, we created a webpage that makes author and expert keywords and related papers search- and browsable: `http://www.keyvis.org/`. Visitors can search all 2629 unique author-assigned keywords, find out which keywords co-occurred how frequently, which manual expert clusters they belong to, and the actual research papers they appear on. Our main goal was to generate an easy-to-use, lightweight interface to our keyword data in order to: (a) support visualization researchers in making more informed decisions when picking keywords for their papers, and (b) give a new lens on identifying relevant related work. We have used the site ourselves for choosing keywords and finding related work that we were not aware of before. We hope that others will find it similarly useful. In the long run, our goal is to maintain the website as a platform for visualization keyword access and analysis.

6 Next Steps

Our analysis is a first step towards two larger research goals:

Creating a common vocabulary: One can think about the problem of creating a common vocabulary for visualization more broadly. By identifying key terms and providing clear definitions, sub-communities in visualization may be able to communicate more clearly about similar approaches and, this, in turn can help to also collaborate more effectively with people outside the community. Finally, a common vocabulary can also facilitate to more easily understand emerging and declining research trends within the field.

Establish a comprehensive taxonomy of visualization research: Keyword analysis could be a first step to develop a taxonomy of visualization research. This could serve two purposes. One the one hand, a taxonomy will help to better communicate "what is visualization" to other disciplines, i. e., researchers and practitioners not part of the VIS community. On the other hand, we are hoping to be able to facilitate the crucial step of matching reviewers with papers and grants such that the peer review process improves and new contributions are seen in the right context.

References

[1] M. Callon, J. Courtial, and F. Laville. Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. *Scientometrics*, 22(1):155–205, 1991.

[2] M. Callon, J. Law, and A. Rip, editors. *Mapping the Dynamics of Science and Technology*. Palgrave Macmillan, Hampshire, UK, 1986.

[3] J. Law, S. Bauin, J.-P. Courtial, and J. Whittaker. Policy and the mapping of scientific change: A co-word analysis of research into environmental acidification. *Scientometrics*, 14(3–4):251–264, 1988.

[4] Y. Liu, J. Goncalves, D. Ferreira, B. Xiao, S. Hosio, and V. Kostakos. CHI 1994–2013: Mapping two decades of intellectual progress through co-word analysis. In *Proc. CHI*, pages 3553–3562. ACM, 2014.

[5] A. D. Ritzhaupt, M. Stewart, P. Smith, and A. E. Barron. An investigation of distance education in North American research literature using co-word analysis. *International Review of Research in Open and Distance Learning*, 11(1):37–60, 2010.