

RESEARCH ARTICLE

Assisting researchers in bibliographic tasks: A new usable, real-time tool for analyzing bibliographies

Antonina Dattolo  | Marco Corbatto 

SASWEB Research Lab, Department of Mathematics, Computer Science, and Physics, University of Udine, Gorizia

Correspondence

Antonina Dattolo, SASWEB Research Lab, Department of Mathematics, Computer Science, and Physics, University of Udine, Via Santa Chiara, 1-30170 Gorizia, Italy.
Email: antonina.dattolo@uniud.it

Abstract

The amount of scientific papers is growing together with the development of science itself; but, although there is an unprecedented availability of large citation indexes, some daily activities of researchers remain time-consuming and poorly supported. In this paper, we present Visual Bibliographies (VisualBib), a real-time visual platform, designed using a zz-structure-based model for linking metadata and a narrative, visual approach for showing bibliographies. VisualBib represents a usable, advanced, and visual tool, which simplifies the management of bibliographies, supports a core set of bibliographic tasks, and helps researchers during complex analyses on scientific bibliographies. We present the variety of metadata formats and visualization methods, proposing two use case scenarios. The maturity of the system implementation allowed us two studies, for evaluating both the effectiveness of VisualBib in providing answers to specific data analysis tasks and to support experienced users during real-life uses. The results of the evaluation are positive and describe an effective and usable platform.

1 | INTRODUCTION

For researchers, publications represent indispensable references for communicating scientific results, and for knowing the advances in specific research areas. A relevant part of these activities falls in the research field dedicated to literature search and review that involves information seeking, sensemaking, comprehension and communication (Bethard & Jurafsky, 2010; Booth et al., 2016; Borgman & Furner, 2002; Zhang et al., 2008), and requires the retrieval, the exploration, the organization, the evaluation, and the sharing of papers, regrouped in bibliographies. Researchers spend a lot of time in exploring bibliographies (Di Iorio et al., 2015), when writing a paper, when navigating in a new research area, when evaluating papers, when searching for related work; often they start from few seed

papers and apply a backward and forward search for finding interesting papers.

Although large citation indexes, like Scopus, Web of Science (WoS), Google Scholar, provide advanced search mechanisms, they return long list of results, data, and metrics, and poorly support researchers in their usual tasks (Bikakis et al., 2021; Sultanum et al., 2020). Starting from these exigences, research has been developed in many directions with the aim of automatically generating related work sections of writing papers (Chen & Zhuge, 2019); identifying the relationships between metadata and criteria to judge the relevance of scientific papers (Zhang et al., 2021); offering context-aware citation recommendations (Jeong et al., 2020); creating interactive storyboard for exploring visual information in scientific publications (Zeng et al., 2020); realizing visual tools for analyzing

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2021 The Authors. *Journal of the Association for Information Science and Technology* published by Wiley Periodicals LLC on behalf of Association for Information Science and Technology.

citation networks (Van Eck & Waltman, 2014) or exploring bibliographies (Federico et al., 2017).

In this context, we have designed and realized Visual Bibliographies (VisualBib), a Web platform, conceived with one main aim: Provide researchers with a real-time visual analytics tool for complex analyses on scientific bibliographies, capable of supporting them during their daily activities and bibliographic tasks (BTs).

We presented a first conceptual model of VisualBib in (Dattolo & Corbato, 2019), where we elaborated an initial visual interface tied at the pure visualization of bibliographic data. On the basis of the results obtained by that study, where we evaluated some usability aspects of the proposed interface (compared to Scopus), we have designed and developed a more articulated platform, which we present in this paper, where we completely have reformulated the original proposal.

Starting from the cognitive document use model, proposed in Wang and Soergel (1998), we have generalized the model, that continues to describe the selection process of papers during activities of literature search and review, but also now includes more general situations, like, for example, the evaluation of the scientific production of a researcher. We have introduced two new stages, and specialized the typologies of available metadata, with the aim to offer *valuable tools* for collecting advanced metadata, and exploring and analyzing a bibliography. In order to reach these objectives, we have analyzed and organized useful *typologies of metadata* (among that exposed by popular bibliographic indexes), and a set of 13 *core BTs*, which we have posed at the basis of the design of our platform; for modeling data and navigation, we have used a *zz-structure-based model*, which offers a *semantic and contextual* added value, and for realizing the visual interface, the exploration and analysis mechanisms, we have employed graphical organizers.

The final result is VisualBib, a Web-based, usable, visual tool assisting researchers in real-time to accomplish a set of core BTs. It supports the creation of bibliographies, based on citation networks; their visualization in a narrative format; the automatic real-time seeking of metadata related to papers and authors and the checking of their consistency, together with a wide set of other advanced and usable features and functions for their exploration and visual analysis.

This paper is organized as follows: In Section 2, we discuss related work; in Section 3, we describe VisualBib, starting from the formulation of a new document use model and from the analysis and selection of available metadata; in Section 4, we introduced the underlying semantic data model; in Section 5, we discuss two main features of the tool, through two use case scenarios. Finally, in Section 6, we describe two studies aimed to

evaluate the effectiveness of the main features of VisualBib (both during guided and free tasks), through a three-step evaluation, carried out as two between-subject studies. Conclusion and future work end the paper.

2 | RELATED WORK

A researcher performs specific BTs depending on the role they play: author, reviewer, supervisor, evaluator, or editors (Dong et al., 2019). We have analyzed these tasks from a general and systematic viewpoint that englobes *literature search and review*, and, then, more specifically, we have focused our attention on the *information visualization* aspects.

2.1 | Literature search and review

The BTs often arise from the researcher's involvement in literature search and review activities. A vast scientific literature emphasizes the importance of a systematic and rigorous approach for these activities, in particular, for documenting the literature search process (Vom Brocke et al., 2009), identifying significant methodologies to apply and augment scholarly communication and bibliometrics (Borgman & Furner, 2002), carrying out a successful literature review (Booth et al., 2016), finding ways to improve the use of scientific papers' indexes (Tenopir et al., 2008); learning literature search models from citation behavior (Bethard & Jurafsky, 2010); supporting sensemaking of research literature (Zhang et al., 2008); opportunely selecting and using available metadata (Bethard & Jurafsky, 2010).

In this process, we have carved out a small space, that, to the best of our knowledge, is not completely covered by existing work. We would like to support researchers, during the selection and analysis of relevant metadata related to a bibliography, by providing them with a set of *visual, analytical tools* that can help them for coming to a judgment/choice about the papers in the bibliography.

The cited literature provided us the conceptual framework (Wang & Soergel, 1998) for a new document use model (presented in Section 3.1), and supported us for carrying out the list of core BTs (proposed in Section 3.3). The final result, the realized platform, finds related work in the research field of information visualization.

2.2 | Information visualization

Three surveys (Costagliola et al., 2018; Federico et al., 2017; Kucher & Kerren, 2015) highlight the

emerging importance of using meaningful visual representations for improving the sharing of publications' metadata within scientific communities.

There are several tools which propose graphic representations of bibliographic data and support researchers in analyzing and exploring data and relationships. Unfortunately, the majority of them emerged some years ago, and are no longer under active development, such as PaperLens (Lee et al., 2005), BiblioViz (Shen et al., 2006), CiteWiz (Elmqvist & Tsigas, 2007), PaperCube (Bergström & Atkinson, 2009), Cybis (Costagliola & Fuccella, 2011), Citeology (Matejka et al., 2012), PivotPaths (Dork et al., 2012), and CitNetExplorer (Van Eck & Waltman, 2014).

None of them is a real-time application, in the sense that they work on pre-built datasets, or allow users to upload limited datasets obtained from WoS or other repositories.

The only three active projects, to the best of our knowledge, are CiteSpace III (Chen et al., 2010), VOSviewer (Van Eck & Waltman, 2010; van Eck & Waltman, 2017), and PaperPoles (He et al., 2019). VOSviewer enables users to visualize bibliometric networks, for example, co-authorship, bibliographic coupling, and co-citation networks, and to tune layout and clustering results, using various parameters. It works on larger networks than VisualBib, and it has been conceived for analyzing entire networks, visually aggregating data, rather than giving relevance to the role of individual papers, authors, and related metadata in a bibliography. VOSviewer does not allow users to automatically create the citation network neither offers tools for supporting refinements on a bibliography. Users may export a bibliography using a proprietary VOSviewer format, the Graph Modeling Language format and the Pajek network format.

CiteSpace facilitates a systematic review of a knowledge domain and reveals how a field of research has evolved, and what topics have attracted attention. Similarly, to VOSviewer, it does not offer tools for supporting refinements on a bibliography, neither for avoiding the duplication of authors. It enables users to export bibliography in WoS format, or in Pajek.

Both the systems offer a complex interface, and are not focused on the exploration, creation, and refinement of bibliographies.

PaperPoles uses citation links to create clusters of related papers. It queries Microsoft Academic Graph (MAG) as real-time data source, using the Academic Knowledge Application Programming Interface (API). The exploratory search mechanism, based on citations, is similar to that used in VisualBib, but PaperPoles does not offer opportunities for visualizing, refining, and exporting a bibliography. The interface is not intuitive.

In sum, the active projects are few, none of them uses “live” repositories (except PaperPoles that uniquely uses MAG), few tools enable users to save and share their bibliographies, the interfaces are not always usable. We propose a Web application that overcomes these limitations; it effectively supports users in the creation, analysis, and sharing of bibliographies of small dimensions; its interface is simple and usable as confirmed by the several evaluation studies.

3 | VISUALBIB

The selection and evaluation of papers to consider for creating or approving a valid and accurate bibliography is a complex process. In order to support researchers in these activities, we propose a new *model of document use*, which specializes the role of a variety of metadata; these, as discussed in (Zhang et al., 2021), stimulate clues, which represent the cognitive stimuli for the formulation of relevance criteria for researchers, when they select a paper as worth for reading and citing.

3.1 | Proposing an extended model of document use

Starting from the model of document use, proposed in Wang and Soergel (1998), we inherit and extend the concept of document information elements (DIEs), defining two new categories of metadata:

- DIEs contain typical metadata of a paper, like title, list of authors, abstract, type, etc. (Wang & Soergel, 1998);
- Author information elements (AIEs) contain metadata related to authors, like name and surname, affiliation, h-index, list of papers, list of subject area, number of citations;
- Bibliography information elements (BIEs) contain the previous metadata in aggregated formats, and represent the basis for contextual, analytics, and visual hints. Example of BIEs are the temporal distribution of papers in a bibliography, its citation and collaboration networks, the distribution and frequencies of publication types (i.e., journals, conferences or books), keywords, and subject areas, the distribution of metrics for groups of papers and authors, such as h-indices, or number of citations.

In Figure 1, we represent our model of document use, which, starting from that proposed in Wang and Soergel (1998) (dashed boxes), contains two new major stages (solid boxes): “Collecting metadata,” and

“Exploring&Analyzing.” We identified five major decision points (green boxes), along the stages of document seeking and use during a research project.

The entire process starts from one (or more) seed papers or authors, derived from an information retrieval process; a researcher may select or definitely discard a paper. In case of selection, both paper and author metadata (resp. DIES and AIEs) are retrieved and aggregated metadata (BIEs) is generated. On this basis, researchers could decide to have received sufficient stimuli to apply relevance criteria for deciding to read and eventually keep in the bibliography that paper, citing it. Otherwise, they may decide to continue to the next stage, where they may explore metadata and views on them and use visual analytics tool for choosing to express a relevance judgment (for example, evaluate the production of an author) or finally consider to have sufficient information for continuing to the next stages and deciding if reading, citing and so keeping the paper in the bibliography.

3.2 | What metadata and what indexing platforms?

In Bethard and Jurafsky (2010), the authors argue that the complexity of scientific literature search can be reduced by building a retrieval model that integrates the features of classic keyword search with features important to scientific literature retrieval, such as the citation network, the recentness of publication, the abstract of an article, and the list of references. These metadata

represented stimulating insights for the selection of metadata to use and visually represent in our system.

In order to define a valid list of DIES and AIEs metadata and derived aggregated BIEs, we analyzed the available data, exposed from a set of popular bibliographic indexes, and we listed them in Figure 2. The richest data sources, among those analyzed, are represented by the Scopus (column A) and WoS (column D) citation indexes. Scopus offers a rich API also available to nonsubscriber users with some restrictions on the accessible fields, on the number of queries per week and the amount of results per call. WoS offers a limited API service, called WoS API Lite, to retrieve basic paper and authors metadata, while a more complete service, WoS API Extended, is available to subscribers only. We decided to interface VisualBib with Scopus, OpenCitations and CrossRef/ Orcid API services, and we used the papers' and authors' metadata, contained in Figure 2 and made available by them, to define the DIES and AIEs for the implementation of VisualBib. Currently VisualBib interfaces, in real-time, with the API services provided by these four citation indexes. We effectively used this characterization of available metadata for implementing VisualBib.

3.3 | Identifying a core set of BTs

Starting from the introduced concepts of DIES, AIEs, and BIEs, we identified the core set of BTs relevant for researchers, shown in Figure 3, in part deriving them by the current literature (Bethard & Jurafsky, 2010; Booth

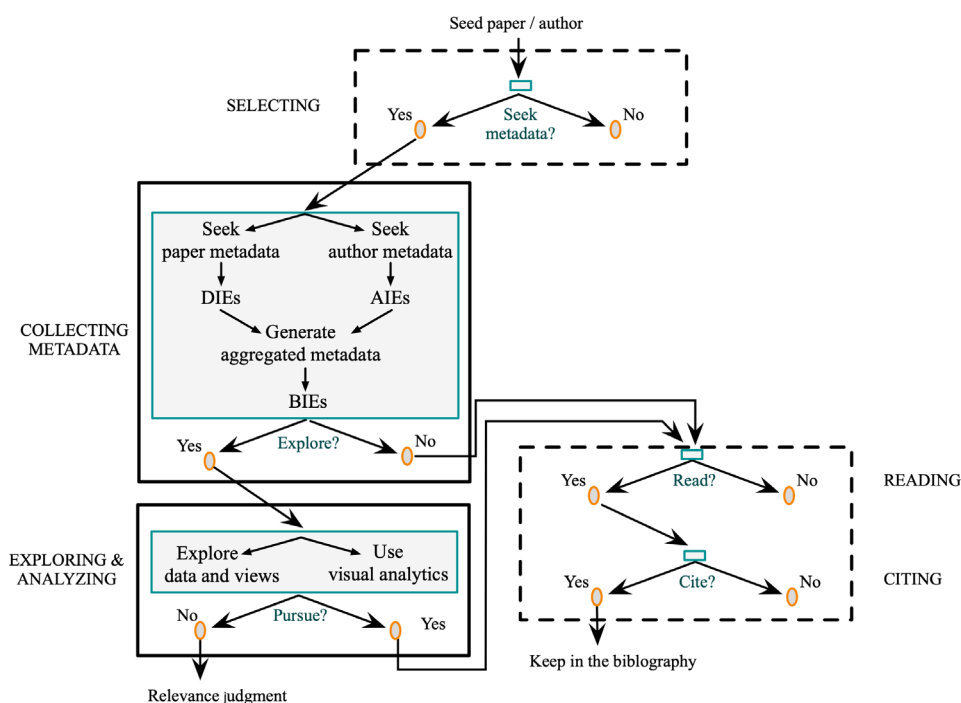


FIGURE 1 The extended model of document use. We specialized document information elements, and added author information elements, and bibliography information elements metadata [Color figure can be viewed at wileyonlinelibrary.com]

FIGURE 2 Comparison of the papers' and authors' data, exposed by 10 bibliographic indexes on the respective portals (not necessarily accessible by application programming interface [API]). A: Scopus, B: Open Citations, C: CrossRef/Orcid, D: WoS, E: OpenAire, F: Google Scholar, G: CiteSeerX, H: Microsoft Academic, I: DBLP, L: AMiner. The A(Scopus) column also shows information on the accessibility of the data via API

Papers' Data	A	B	C	D	E	F	G	H	I	L
Title & pub. year	●/●	●	●	●	●	●	●	●	●	●
Abstract	●/●	○	●	●	●	●	●	●	○	●
List of authors	●/●	●	●	●	●	●	●	●	●	●
Doc. type	●/●	●	●	●	○	○	○	○	○	○
OpenAccess Flag	●/●	●	○	●	●	○	○	○	○	○
# of citations	●/●	●	●	●	○	●	●	●	○	●
Citation benchmarks	●/○	○	○	●	○	○	○	○	○	○
PlumX metrics	●/○	○	○	○	○	○	○	○	○	○
Authors' keywords	●/●	○	○	●	○	○	●	●	○	○
Indexed keywords*	●/●	○	○	●	○	○	○	●	○	○
Topic list**	●/●	○	○	●	○	○	○	●	○	○
Journal title	●/●	○	●	●	○	○	○	●	○	●
Event title	○/○	○	●	○	○	○	○	○	●	○
# & list of references	●/●	●	●	●	○	○	●	●	○	●
# & list of citations	●/●	●	○	●	○	●	●	●	○	●
Related papers	●/○	○	○	○	○	●	●	●	○	●
Docs sharing refs	○/○	○	○	●	○	○	●	○	○	○
Authors' Data										
List of papers	●/●	●	●	●	○	●	●	●	●	●
List of subjects**	●/●	○	○	●	○	●	○	●	○	●
# of citations	●/●	○	○	●	○	○	○	●	○	●
List of citing docs	●/○	○	○	●	○	○	○	●	○	○
List of all cited refs	●/○	○	○	○	○	○	○	○	○	○
List of co-authors	●/○	○	○	●	○	●	○	●	●	●
H-index	●/●	○	○	●	○	●	○	○	○	●
Doc dist. by source	●/○	○	○	●	○	○	○	●	○	○
Doc dist. by type	●/○	○	○	●	○	○	○	○	●	○
Doc dist. by sub. area	●/●	○	○	●	○	○	○	○	○	●
History***	●/●	○	○	●	○	○	○	○	○	●

*keywords provided by API are those listed as “Engineering uncontrolled terms” in the Scopus document page. A weight is associated to each keyword.

**topics and subject-areas are provided with a description, an abbreviation and a code, see <https://api.elsevier.com/content/subject/scopus>.

***both affiliations and journal publication history are provided.

FIGURE 3 The core set of bibliographic tasks used for designing and realizing VisualBib

- BT1: Retrieve papers and authors metadata by querying citation indexes.
- BT2: Visualize metadata.
- BT3: Annotate papers.
- BT4: Visualize the temporal distribution of set of papers.
- BT5: Distinguish among the different typologies of papers.
- BT6: Identify subject areas and keywords.
- BT7: Identify the collaboration network between authors in a bibliography.
- BT8: Investigate the citation network in order to find new related work.
- BT9: Highlight and correct data inconsistencies.
- BT10: Filter papers and authors applying different ordering criteria.
- BT11: Introduce tags for identifying cluster of papers for different aims.
- BT12: Compare metrics associated to subsets of papers and/or in a bibliography.
- BT13: Import, export, save, and share a bibliography.

et al., 2016; Borgman & Furner, 2002; Dong et al., 2019; Elmquist & Tsigas, 2007; Ponsard et al., 2016; Tenopir et al., 2008; Vom Brocke et al., 2009; Zhang et al., 2008; Zhang et al., 2021). These tasks have guided the realization of VisualBib. Conceptually, we identify four macro-categories of tasks: *Manage a bibliography*, that identifies basic tasks, related to import metadata from citation indexes or datasets, export, save, and share a bibliography

(BT1 and BT13); *visualize and explore* a bibliography (BT2, BT4-BT7), that supports users in exploring the metadata through a visual interface; *analyze and compare* (BT8-B10, BT12) enables users to apply appropriate filters, perform complex analytics, and apply backward/forward search mechanisms for identifying new interesting papers; finally, *annotate and tag* (BT3, BT11) proposes a way to annotate individual papers and semantically tag groups of them. In

next Section 3.4, we will describe VisualBib associating its features to the core set of BTs.

3.4 | Our proposal

The aim of VisualBib is to target analytical challenges in BTs, supporting researchers by means of a real-time visual platform, which provides users with some tools to conduct complex analyses.

VisualBib, freely available at <http://visualbib.uniud.it>, is organized as a single page Web application, based on W3C standard languages, such as HTML5, CSS3, and SVG; it adopts the D3js (Bostock et al., 2011) framework for data and DOM manipulation, and the visual element management; it uses AJAX techniques to perform Cross Origin Resource Sharing) calls and client-server interactions. VisualBib is compatible with modern browsers; known incompatibilities have been detected with Mozilla Firefox version older than 51, Microsoft Explorer and Safari. VisualBib notifies the user in case of an incompatible browser. Since VisualBib queries the Scopus APIs and Scopus is a commercial service, the user must navigate from a subscriber's domain; this restriction does not apply to the other used indexes, OpenCitations and CrossRef/Orcid, being open access.

The graphical interface of VisualBib is organized in five main environments, enclosed in Figure 4 in dashed boxes.

3.4.1 | Command, narrative view, and metadata panels

Different from the major part of bibliographic management and analytical tools, VisualBib queries in real-time Scopus, OpenCitations, and/or CrossRef/Orcid (BT1), and may upload specific papers or a whole bibliography (for example, the scientific production of an author), but also seek detailed DIES, AIEs and BIEs, or exploring and extending the current set of papers to cited and citing papers. In Figure 4, the uploaded bibliography is visualized in the central panel, in the format of a narrative view. A part of the rich set of DIES, AIEs and basic BIEs is visualized (Figure 4-bottom) in the Metadata panel (BT2), and exploded in more detail in Figure 5, where it is visible that any paper can be annotated by compiling the "User's notes" field (BT3).

The narrative view, the exploration and the analysis panels aggregate these metadata in order to give them a visual representation (*advanced BT2*), and provide useful functionalities for supporting the analytical study of a bibliography. While the command panel supports basic

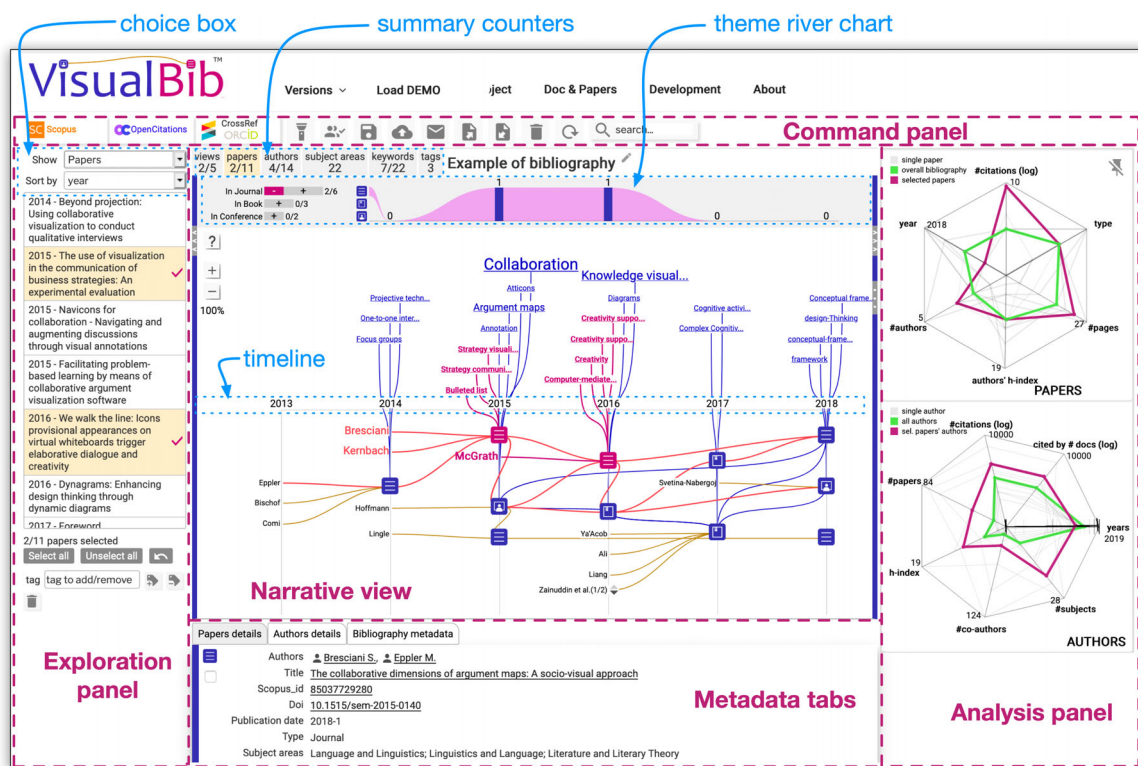


FIGURE 4 The graphical layout of VisualBib. The shown narrative view is reachable clicking on "Load Demo," and then on "Subject areas" in the summary counters [Color figure can be viewed at wileyonlinelibrary.com]

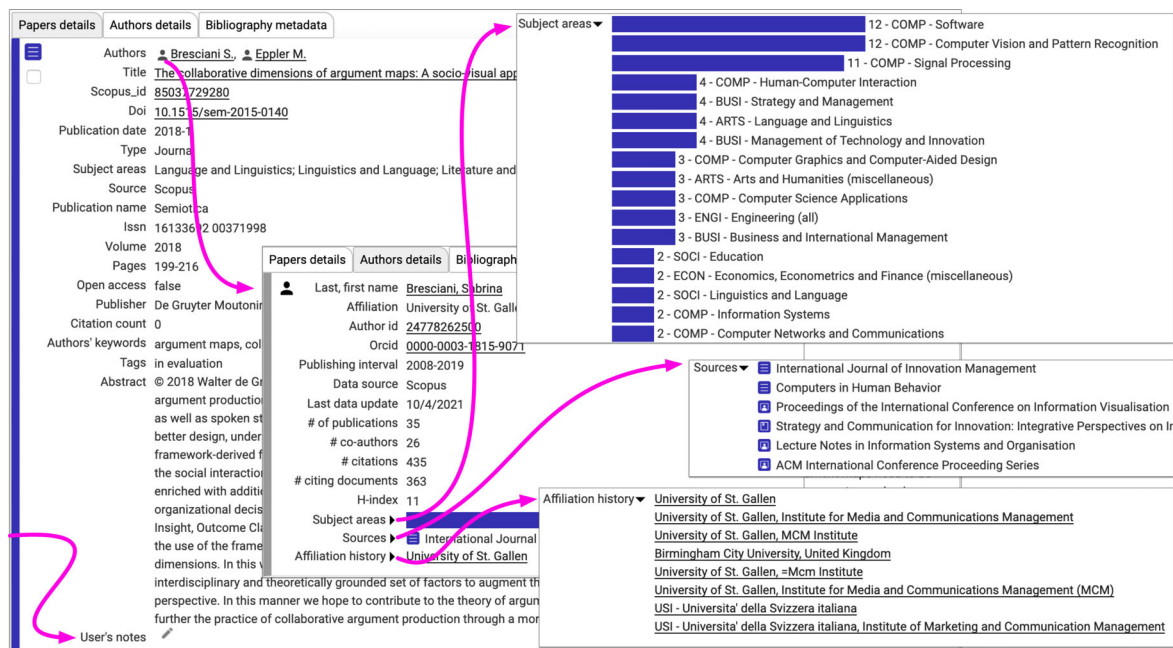


FIGURE 5 Document information elements and author information elements in the metadata panel. The author icons, on the left of authors' names, allow users to open the corresponding metadata on the second tab; the triangle symbols to visualize subject areas, sources, and affiliation history of an author; finally, papers' titles and authors are linked to external pages related to the source of the data [Color figure can be viewed at wileyonlinelibrary.com]

and essential tasks, such as the import of metadata from citation indexes or datasets, the import/export from/to other formats (like BibTeX), the saving and sharing operation on a bibliography (BT1 and BT13), the narrative view is dedicated to visualization and exploration (BT2, BT4-BT8); it presents a bibliography, graphically arranged in a 2-dimensional space, where the horizontal dimension is the *timeline*, discretized by years, while the vertical dimension is spatial and is used to properly organize authors' names, papers, and their relationships. The diagram includes the last names of the authors involved in at least one paper of the current set: From each author starts a goldenrod line that connects all their papers, from the oldest to the newest. The papers are represented by blue, round-cornered square items, which become magenta when are selected. Moving the mouse over an author, all their papers written are highlighted; similarly, moving the mouse over a paper, all the co-authors are highlighted, together to the papers written by them, also separately (contributing to BT7).

The narrative view panel contains, in addition to the narrative view:

- The summary counters, which act as quick reference for visualizing numbers of views, papers, authors, subject areas, keywords, and tag selected/present in the current bibliography but also for quickly accessing to the choice box of the exploration panel.

- The theme river chart, a clickable histogram, which shows the number of documents for each type (journal, conference or book papers), associated to a related icon (BT5); moving the mouse over the colored areas or on one of the three publication types, the histogram changes for showing the corresponding counters.
- Optional word clouds (BT6 and partially BT11), where it is possible to visualize *subject areas*, *keywords*, or user *tags*. Each term in the word cloud is connected with all the related papers by a path, that can be highlighted, together with the involved papers icons, by moving the mouse over the word cloud.

In Figure 4, the narrative view proposes a bibliography of 11 papers, written by 14 authors between 2014 and 2018, connected by 22 subject areas, 22 keywords, and 3 tags, added from who is editing the bibliography. In the current view, two 2 papers have been selected (in fact, they appear in magenta), and, as the theme river chart shows, they are two journal papers; moreover, “subject areas” view is also present, where it is possible to observe that the most frequent subject areas, in the current bibliography, are “Computer Science” and “Philosophy.”

A click on any paper in the narrative view opens a popup window, containing some DIES, as shown in Figure 6-right: A second click on the four-arrows icon loads, in a separate form, the list of cited/citing papers (Figure 6-left), retrieved from the selected citation index

(Scopus, in the example. See Source: Scopus). This feature enables backward (on older literature cited in the paper) and forward (on additional papers that have cited the current one) search (BT8). Within each selection list, containing respectively all the cited/citing papers, each paper, already in the diagram, appears highlighted in blue. Users may select, from the two lists, the documents of their interest and upload them in the narrative view.

VisualBib enables users also to import a bibliography in BibTeX format: This opportunity may generate some inconsistencies on metadata, due to the different citation indexes, that use not-shared identifiers for papers and authors; or due to misspellings in names or to the presence of special characters, or in general to incorrect information. Some inconsistencies, like on year, order of authors, number of pages may be automatically corrected using the function “Seek metadata” that analyzes the set of papers and authors of the current bibliography, checking, for each of them, the availability of extended metadata through the abstract retrieval and author retrieval Scopus APIs. The “Seek metadata” (BT1) can be applied to any paper and related authors (by clicking the “torch” icon in the popup window—see in Figure 6) or to the entire bibliography (by clicking the “torch” icon on the command panel).

During the analysis of a paper, the list of authors retrieved from Scopus is compared with the current authors' list of the papers in the bibliography: Authors uniquely identified are automatically matched, otherwise a textual comparison is performed; if the matching is not exact, a new author is introduced and associated to the paper. This procedure covers also the case of an incomplete author list.

Possible duplication of authors will require the “Match authors” wizard (BT9) that offers the opportunity to disambiguate authors' names, avoid duplications of same authors, and correct errors, providing a tool for overcoming typical author name disambiguation (Kim & Kim, 2020).

3.4.2 | Exploration panel

The exploration panel contributes to support BT2, BT5-BT8, BT10-BT11, and enables users to explore, by means of the *choice box* “Show/Sort by” the lists of papers, authors, subject areas, keywords, and tags of the bibliography (BT5-BT8), to change the sorting criteria (BT10), to activate/deactivate the supported views (the views present aggregated metadata—BT2), to display the frequency distributions of the items (BT4-BT5), to select/deselect papers and to apply/remove *personal tags* to papers (BT11). Figure 7 explodes the tabs of this panel: The *views tab* allows users to show/hide five typologies of connections between papers: authors, citations, subject areas, keywords, and user-added tags. In Figure 4, are active two views: Authors and Keywords, while an example of citations view is visible in Figure 14-bottom, in the context of the use case scenario proposed in Section 5.2. The citation view represents a way to visualize the connections between papers, and, together the backward/forward search, supports the realization of BT8.

The *papers tab* lists the papers in the bibliography, which may be displayed, changing the sorting criteria (year, title, number of citations, number of authors and type); moving the mouse over a paper in this list (in the

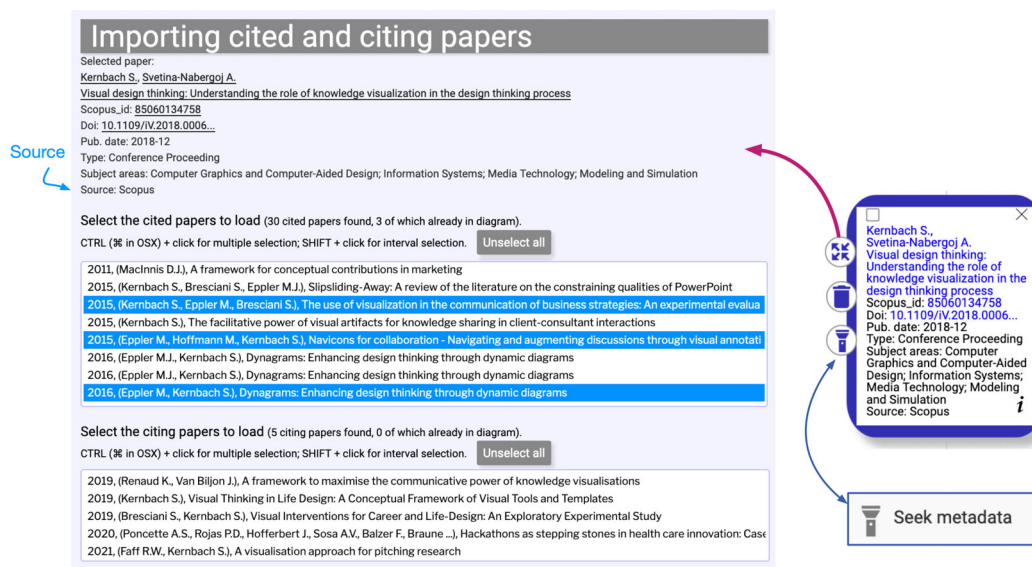


FIGURE 6 A tool for realizing backward and forward search [Color figure can be viewed at wileyonlinelibrary.com]

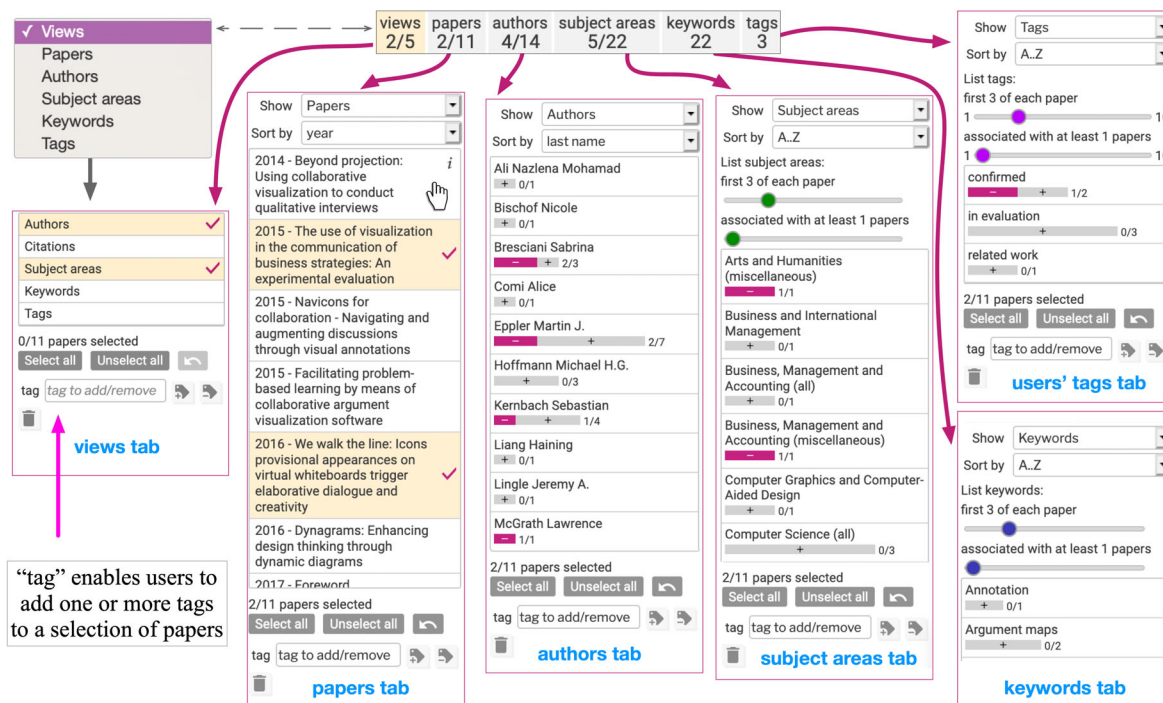


FIGURE 7 The six main tabs of the exploration panel [Color figure can be viewed at wileyonlinelibrary.com]

example of Figure 7: “2014—Beyond projection: Using...” appears the “i” icon, which enables users to inspect the detailed metadata tab. The *other four tabs* visualize respectively the list of authors, subject areas, keywords, and tags. Each tab has its own list of sorting criteria and displays a bar chart representing the absolute frequencies of each item in the current bibliography. The bars are marked with a clickable “+” icon for adding new papers to the current selection. The mechanisms of papers selection, by clicking on authors, subject areas, keywords, or tags, facilitates the visual analysis.

An interesting opportunity for user is to apply/remove personal tags. This offers a way to *semantically group papers* according to different criteria: For example, users can specify a class for some papers (e.g., related work, reference to a theory/method/tool, etc.) and/or to mark a status (in evaluation or confirmed paper, etc.) and/or a relevance (highly significant paper or secondary reference, etc.), or to explicit the rationale of the paper according to a taxonomy of classes (Background, Motivation, Uses, Extension, Comparison, or Contrast, Future) (Jurgens et al., 2018).

The user interaction on each of these environments propagates its effect to the other environments, maintaining them synchronized; for example, the selection of the two papers (2015—The use of visualization...; 2016—We walk the line...) in the exploration panel of Figure 4 (and Figure 7) leads to visual changes and appropriate highlights in the summary counters (views 2/5, papers 2/11, etc.), in the theme river chart (2/6 selected journal papers), in the

narrative view (the two papers are highlighted in purple) and in the “selected papers” of the analysis panel.

3.4.3 | Analysis panel

The analysis panel (Figure 4-right) contains two radar charts, the first is related to papers metrics, while the second is related to authors metrics; it supports *BT2*, *BT10*, and *BT12*. Both the radar charts include a path for each paper (resp. author) in the bibliography (gray lines) and a path representing the median values of the metrics for the overall bibliography (in green) (*BT2*).

If a user selects (*BT10*) some papers of the bibliography, the radar presents also the aggregate path (in magenta), based on the medians. If the number of selected papers is exactly 2, for facilitating the comparison, the radar displays also two additional paths related to the selected papers, highlighted respectively in orange and azure colors (*BT12*). In order to highlight the path of a specific paper and visualize the values of its metrics, users can move the mouse over the corresponding item in the exploration panel or over its paper icon in the narrative view. It is also possible to interact with the chart, moving the mouse over the specific paths: In this case, the paths are highlighted in black and the metric values are displayed. These features enable users to easily point out papers with particular values for a metric: The title of the paper appears and the corresponding icon in the narrative view is highlighted.

4 | USING SEMANTIC STRUCTURE FOR MODELING DATA AND NAVIGATION

Narrative views, citation networks, word clouds for subject areas, keywords, and tags, and all the data organization and their visualization take advantages from the underlying data model, which specializes *zz*-structures for semantic linking papers, authors, and related metadata.

A *zz*-structure can be thought as a space filled with cells, called *zz*-cells, connected into linear sequences. Cells are connected together with links of the same color (corresponding to the same semantic relationship) into linear sequences, called *dimensions*. We compared them with potential candidates, such as resource description framework (RDF) or JavaScript Object Notation (JSON):

- In the level of graph data model, the underlying structure of any RDF expression is a directed labeled graph (or multigraph), which consists of nodes and labeled directed arcs that link pairs of nodes; for JSON, is an edge-labeled tree (Bourhis et al., 2017). As proved in (McGuffin & Schraefel, 2004), *zz*-structures are general data structures, since subsume lists, 2D arrays, trees, polyarchies, and are equivalent to edge-colored directed multigraphs.
- At conceptual level, RDF is based on an assertional logic, in which triple expresses a proposition. This imposes a fairly strict monotonic, limiting its expressive power; in their essence, JSON documents are dictionaries consisting of key-value pairs, where the value can again be a JSON document, thus allowing an arbitrary level of nesting. Instead, *zz*-structures introduce an intrinsically nonhierarchical, node-centric system of conventions for data and computing (Dattolo & Luccio, 2009a; Dattolo & Luccio, 2009b; Nelson, 2004), which emphasizes the expressive power, and allows the use of non-monothonic constructs.

In this section, we focus our attention on the exploration panel, and we provide for it a formal definition in terms of *zz*-structure.

Definition. The exploration panel *EP* is as a *zz*-structure, where

- $V = \{P, A, SA, K, T\}$, the finite set of vertices is composed by finite sets of papers: resp. authors, subject areas, keywords, and tags;
- $D = \{Views, Papers, Authors, SubAreas, Keywords, Tags\}$ are sets of dimensions.

Each set (except Views) contains all the items of the corresponding set of vertices, which connect following a sorting criterion.

- $Views = \{V\text{-authors}, V\text{-citations}, V\text{-SubjectAreas}, V\text{-Keywords}, V\text{-Tags}\}$;
- $Papers = \{P\text{-year}, P\text{-title}, P\text{-\#citations}, P\text{-\#authors}, P\text{-type}\}$;
- $Authors = \{A\text{-lastname}, A\text{-hindex}, A\text{-\#citations}, A\text{-\#papers}, A\text{-\#papers-in}, A\text{-\#co-authors}\}$;
- $SubAreas = \{SA\text{-\#a..z}, SA\text{-\#occurrences}, SA\text{-firstposition}\}$;
- $Keywords = \{K\text{-\#a..z}, K\text{-\#occurrences}, K\text{-firstposition}, K\text{-author-system}\}$

$Tags = \{T\text{-\#a..z} \quad T\text{-\#occurrences} \quad T\text{-firstposition}\}$

This definition describes the static data model of VisualBib, which becomes dynamic applying a set of views and other mechanisms for the visualization and navigation.

4.1 | *zz*-views

zz-structures generate a pseudo-space that is somewhat comprehensible visually, although there is not a unique canonical viewing mechanism (Nelson, 2004). In any case, views for a *zz*-structure are called *zz*-views. At the moment, VisualBib supports five possible *zz*-views; they are specified in the dimension *Views* and are: the *narrative zz-view* (formally defined in Dattolo & Corbato, 2019); the *citations zz-view*, shown in Figure 15-bottom, and the three *word cloud zz-views*, related to subject areas (shown in Figure 4), keywords and tags.

Figure 8 provides a graph representing a fragment of the *zz*-structure *EP*; the involved dimensions are *Views*, which proposes the five possible *zz*-views, available in VisualBib; *P-year*, which contains the papers, sorted by year of publication; *A-lastname* and *SA-a.....z* by alphabetic order resp. of last names and of subject areas; *K-firstposition* by the position occupied by keywords in the papers, and *T-\#occurrences*, by the most used tags.

Each dimension is characterized by a color (=type of line), which is associated to a semantic label; when, like in this case, we would like to graphically emphasize the role of a dimension, the dimension may become a special *zz*-cell, called *maincell*: A main cell stands for the whole, and connects directly to its supporting cells. All the cells contained in the first row of Figure 8 are main cells. Also, in the example of Figure 9, the cells related to the three

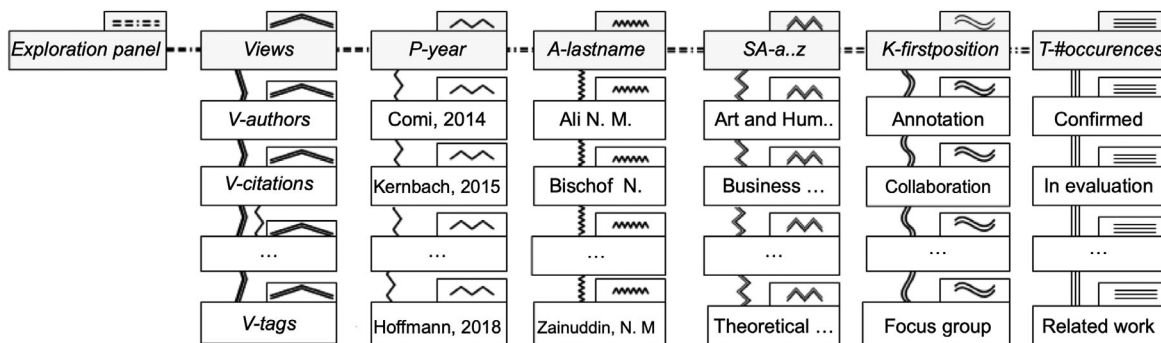


FIGURE 8 A fragment of the zz-structure of “exploration panel” (Figure 7) for the bibliography of Figure 4

authors (a_1 , a_2 , and a_3) are maincells, and we represented them using the same color of their dimensions.

In Figure 9, we consider a fragment of the narrative view in Figure 4, where we labeled the papers with $p_1...p_8$. Using red lines, (resp. solid, dashed and dotted), we identify the authorship dimensions related to three authors (Bresciani, Eppler, Kernbak), and in dotted-blue the dimension related to the subject area (Computer Science). The papers p_2, p_6, p_7, p_8 have co-author Kernbak, p_2, p_3, p_5 Bresciani, while all the eight papers Eppler; only two papers p_4 and p_6 share the common subject area “Computer Science.” These four dimensions represent semantic interconnections between papers; starting from this set of papers and authors would be possible to enrich the graph, visualizing all the other possible dimensions; all the contextual dimensions may be highlighted on demand, opening a holistic view.

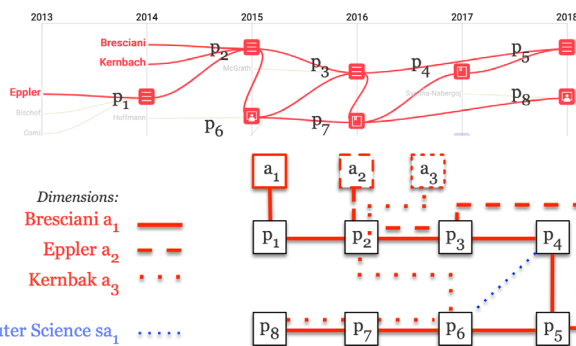


FIGURE 9 A fragment of the narrative view extracted from Figure 4 and the zz-structure related to the eight papers, using four dimensions: “Bresciani,” “Eppler,” and “Kernbak,” and “Computer Science” [Color figure can be viewed at wileyonlinelibrary.com]

5 | USE CASE SCENARIOS

In this section, we present two use case scenarios, focused on two significant aspects of our tool: the variety of metadata formats, and the variety of visualization methods, with the aim to emphasize their usefulness and importance in finalizing BTs. Interested readers will find the links to visualize two video demonstrations related to two additional scenarios in Appendix S1.

5.1 | Scenario 1: The role of metadata

The first scenario shows some typical operations in construction and analysis of a bibliography: Importing a BibTeX archive, retrieving metadata of the papers (DIEs) and involved authors (AIEs), computing, and visualizing aggregated metadata (BIEs) such as citation and collaboration network, frequency and distributions of keywords, and types of papers.

We decided to import a BibTeX file, containing limited metadata and some incorrect data: 15 references, including only the title of papers, their document type, the publication year, and, if available, their scopusid or doi; a typo in an author name, an incorrect publication year for a paper, and an incorrect order of authors’ names. Figure 10 shows the result of this importation, where we skipped the seek metadata phase proposed by the system (see top-right). The narrative view and the metadata tabs show the poor available metadata: In the narrative view, only eight authors (that specified in the BibTeX) are present and authorial links between the paper icons lack.

The next step, illustrated in Figure 11, is to retrieve extended metadata through *seek metadata*. In the case 0 of Figure 11, authors are uniquely identified and their metadata loaded (and corrected - now the publication year is 2018), while, in the absence of this information (case 1), the provided author names are still visible as metadata of any paper, but not linked to any of their articles.

Once the metadata are retrieved, it will be possible to apply the “match authors” wizard, shown in Figure 12,

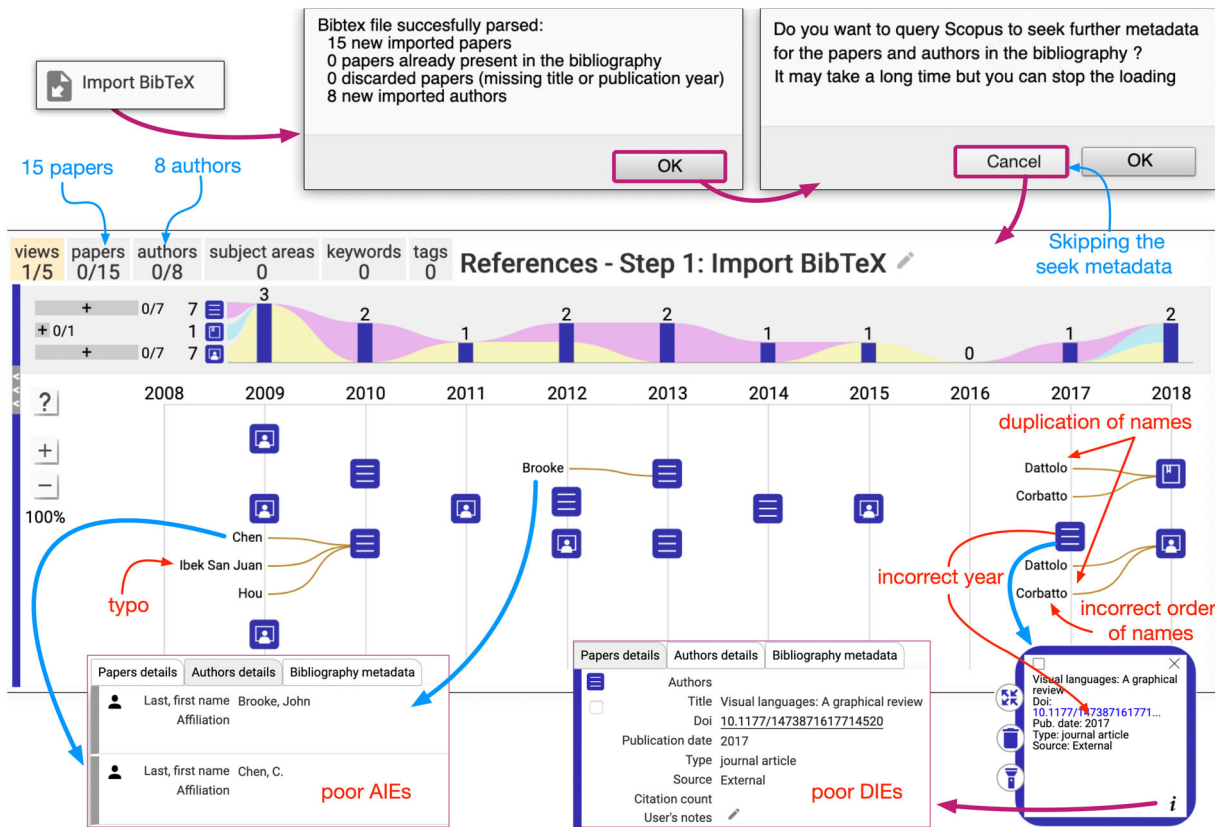


FIGURE 10 In absence of seek metadata, the narrative view, and the metadata tabs lack important visual items and metadata [Color figure can be viewed at wileyonlinelibrary.com]

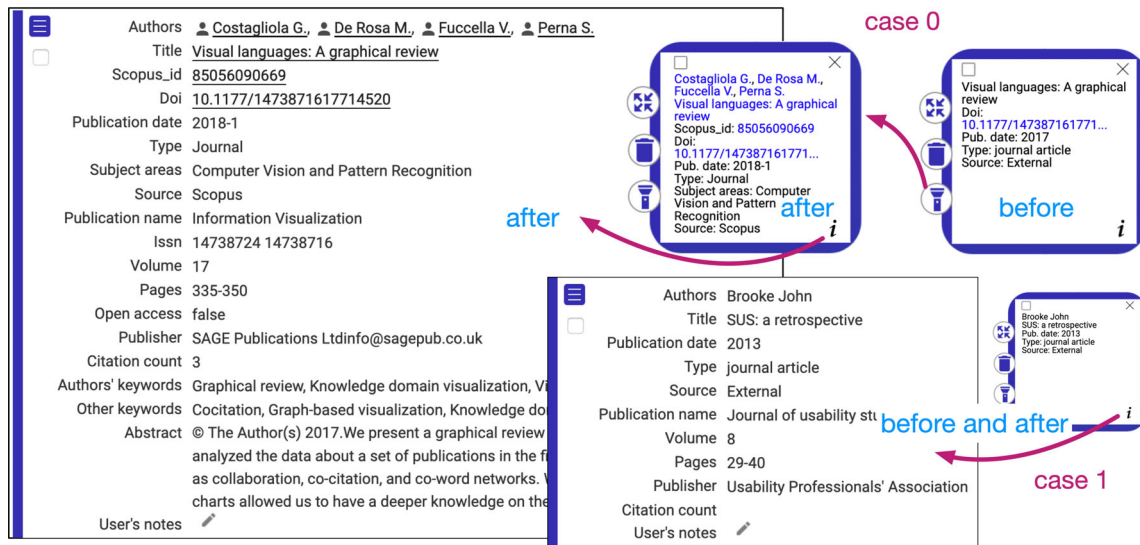


FIGURE 11 Case 0: after the “Seek metadata” the metadata have been enriched. Case 1: no new details [Color figure can be viewed at wileyonlinelibrary.com]

for correcting the duplication of names, the incorrect orders of two authors and the typo.

For the case of duplication and incorrect order, we merge the duplicated (Corbatto)/triplicated (Dattolo) last names, automatically generating also the correct order of

authors. For the case of duplication with a typo, we select the check box, which allows us to examine all the authors, and decide to merge any set of them. Once merged, all the papers associated to each entry will be connected to the unified author entry. Finally, the system

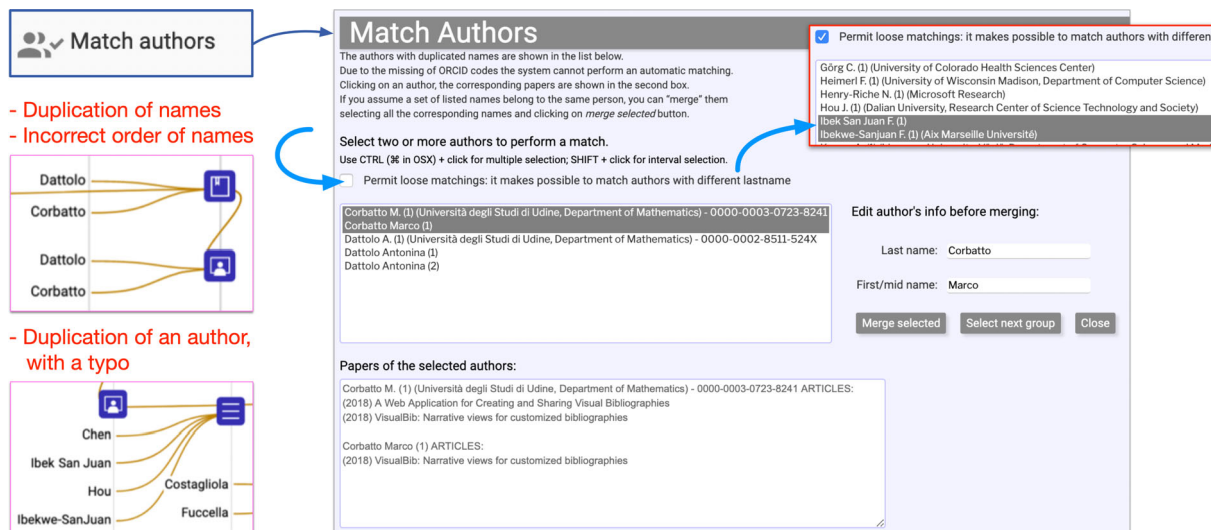


FIGURE 12 The match authors wizard [Color figure can be viewed at wileyonlinelibrary.com]

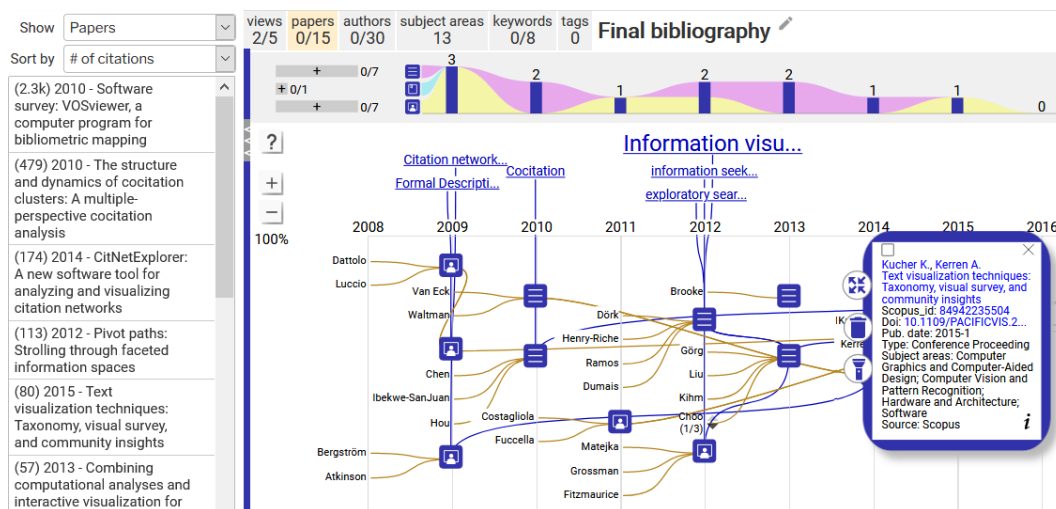


FIGURE 13 Part of the narrative view after the seek metadata phase: It shows the connection between the authors and the papers and some of the filtered keywords (the first two of each paper occurring in at least two papers). On the left, we may see the list of the papers ordered by number of citations [Color figure can be viewed at wileyonlinelibrary.com]

builds the narrative view visible in Figure 13, where all 30 authors are finally included. The view allows users an interactive exploration by moving the mouse over specific items (papers, authors, keywords, ThemeRiver, ...) in order to highlight connected elements along multiple dimensions. It offers users a flexible environment to focus attention on several aspects of the bibliography and, at the same time, to consider the multiple semantic connections of the elements. From this main view, users can perform multiple actions: explore additional candidate papers of a specific author; browse all the references or the citing articles of a certain paper, applying a backward/forward search; in order to expand the bibliography, examine the list of papers, authors, subject areas,

and keywords, ordering them according to different criteria; select a subset of papers in order to semantically tag them or compare their aggregated metrics against the entire bibliography.

5.2 | Scenario 2: The role of visualization methods

The second use case scenario concerns the analysis of the scientific production of an author. We analyze their collaborations, visualize some metrics about their work, and compare them with those of a subset of papers. The narrative view and the theme river chart of Figure 14-top

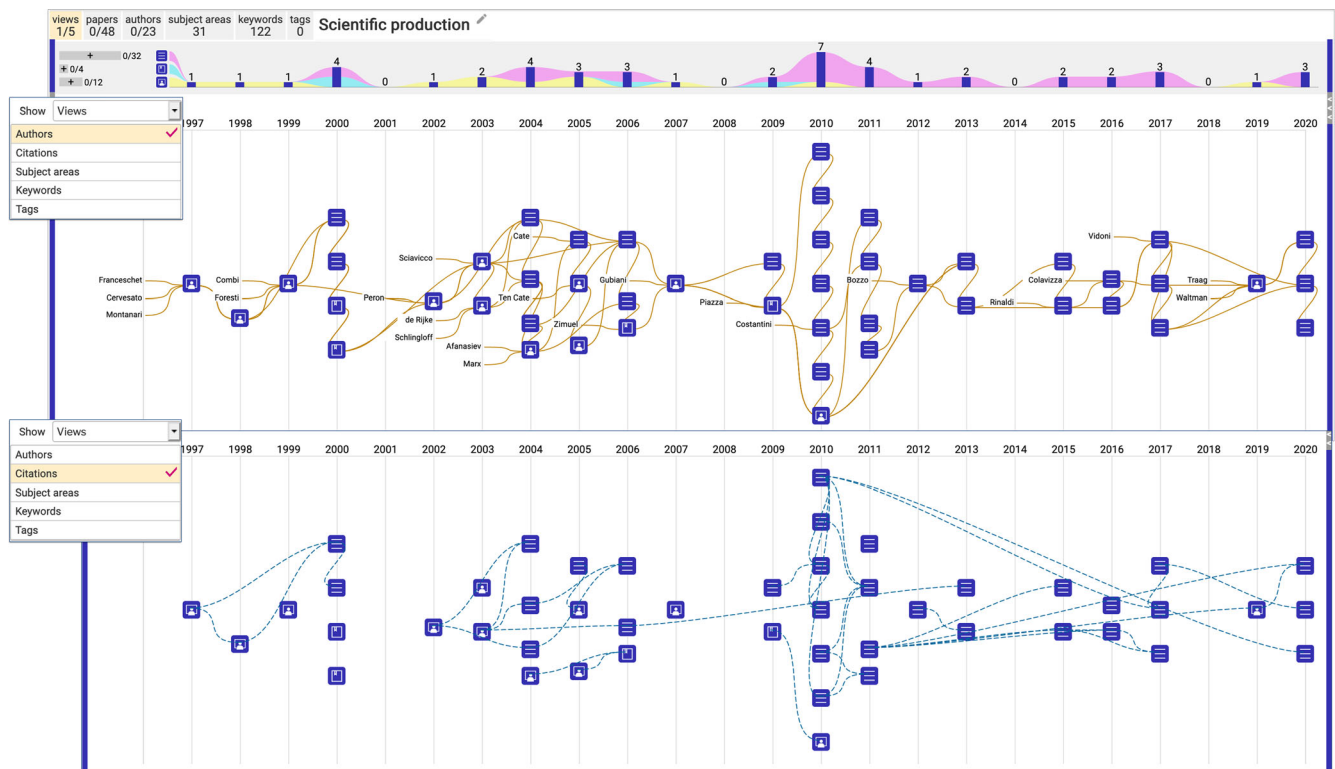


FIGURE 14 The scientific production of an author (top) and their self-citations (bottom) [Color figure can be viewed at wileyonlinelibrary.com]

visually suggest four phases of their research activities, while Figure 14-bottom shows the self-citations. In order to explore the research topics, we may interact with the word clouds of subject areas and keywords, by moving the mouse over each label to highlight related papers as shown in Figure 15-top. The author wrote 48 papers collaborating with 23 co-authors, as indicated by the summary counters: By moving cursor over each author label, all the co-signed papers are highlighted in red. Both occasional and continuous collaborations become evident (Figure 15-bottom), even in quantitative form, as stated in the list of the authors ordered by “# of papers in bibliography.”

By selecting all the papers in common with a specific author (clicking on the corresponding bar in the author list), the *analysis panel* of Figure 16-left enables researchers to compare the aggregated metadata of the selected papers (magenta line) and those of the entire bibliography (green line). By moving the mouse on the lines, we may see the median values for the considered metadata. For example, in this scenario, the median value of number of citations for the selected papers is three in comparison with a median of nine for the entire bibliography; on the other hand, the radar of Figure 16-right states a higher median impact of the authors of the selected papers based on the number of citations (median of 1,086), in comparison with all authors in the bibliography (median 378). This example shows how the system

helps us to conclude that the impact of the joint research of the two authors, placed in the first part of the analyzed author's career, is on average lower than the impact of the entire production but that, in absolute terms, the impact of the two authors is, on average, higher than the aggregate of all contributors. This kind of analysis can also be carried out by considering groups of authors or by clustering papers according to different metadata such as subject areas, keywords, or an arbitrary aggregation made by applying tags.

6 | USER EVALUATION

To evaluate the VisualBib, we carried out two between-subject studies. The main questions that we faced in the analysis of user evaluation are:

1. How difficult is it for users without a specific training on the platform to carry out some analysis tasks on a bibliography?
2. How effectively VisualBib supports different general objectives?
3. Is the system usable according to a standard index such as SUS (System Usability Scale)?
4. How effective do users find the various sections of the application in performing BTs?

FIGURE 15 Interacting with subject area word cloud (top) and the author labels in the narrative view (bottom) [Color figure can be viewed at wileyonlinelibrary.com]

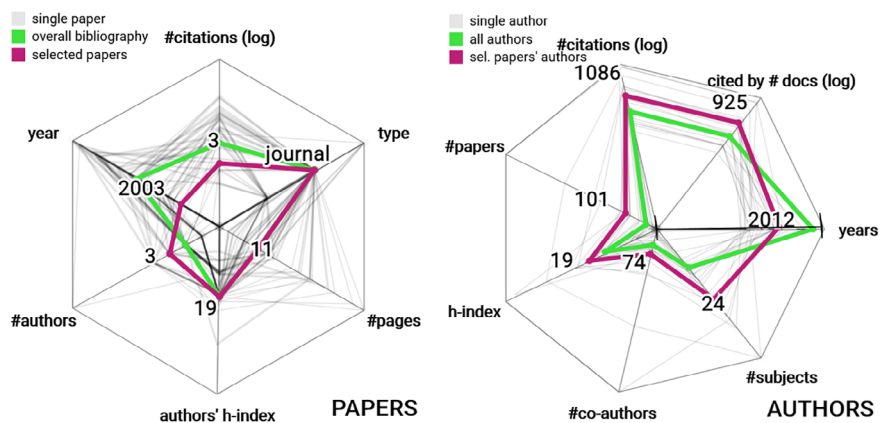
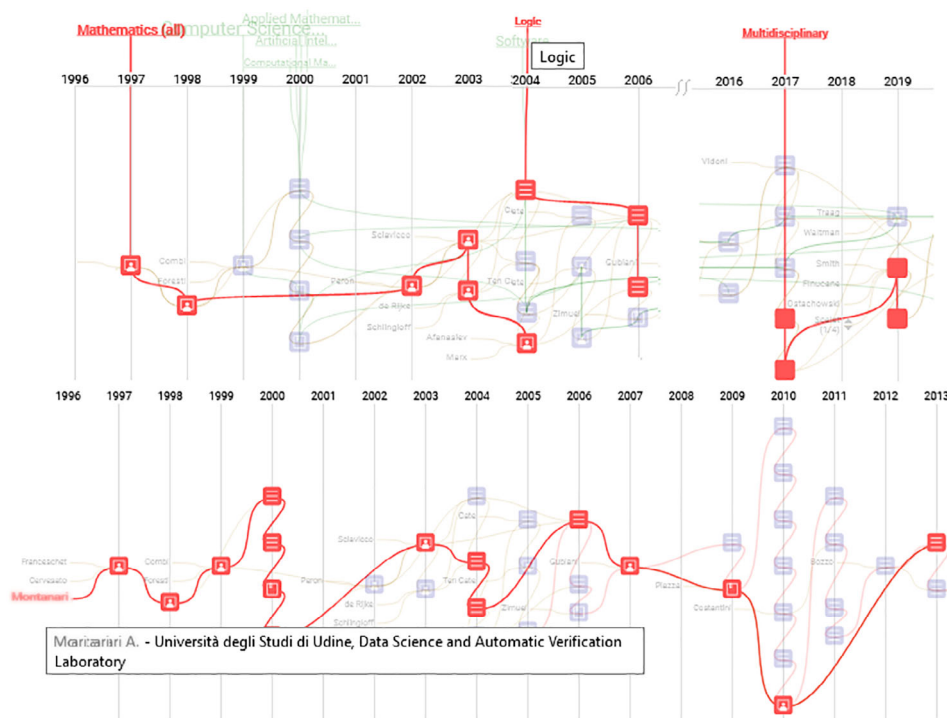


FIGURE 16 The analysis panel: Paper (left) and author (right) radar charts of the bibliography of scenario 2, after selecting the papers in common with the main co-author [Color figure can be viewed at wileyonlinelibrary.com]

6.1 | Participants

We decided to divide the participants in two groups, in two different sessions. All the participants were volunteers who received no compensation and were recruited through personal contact.

Group 1: The first group was characterized by participants from University of Udine (MSc students, librarians, and researchers), familiar with bibliographic research studies, but *unfamiliar* with the use of IT tools to manage bibliographies; it was composed by a sample of 25 participants (16 M, 9 F), age ranged from 20 to 57 ($M = 29.5$, $SD = 11.8$).

Group 2: The second group was characterized by researchers from four European Universities (Spain, UK, Italy), with high level of experience in BTs on scientific literature and familiar with the use of dedicated IT tools;

it was composed by a sample of eight participants (6 M, 2 F), age ranged from 24 to 60 ($M = 40.0$, $SD = 14.6$).

6.2 | Apparatus and procedures

The rationale for the choice of the two groups was to study separately the performance on lower-level tasks for generic users to highlight the effectiveness of the system in providing answers to specific data analysis tasks (*group 1*) and then to collect the opinions on higher level use of the system to highlight its capability to support experienced users (*group 2*).

All the participants were free to choose their favorite operating system and Web browser using a personal computer/laptop/tablet, as long as it is at least 15 in. long. As

Web browser, they used Google Chrome (84%), Mozilla Firefox (12%), and Microsoft Edge (4%). We started with a common live presentation to illustrate the platform and its sections, followed by two differentiated paths.

For *group 1*: An *autonomous exploration* of the platform (20 min); a *training session* (30 min), dedicated to a set of guided tasks: importing a BibTeX archive, retrieving extended metadata from Scopus, matching the author names, integrating the bibliography with additional papers, analyzing the bibliography by extracting some aggregate metadata and finally exporting the enriched bibliography in BibTeX format; a test on *eight quantitative tasks*. Our aim was to measure the difficulty to carry out a set of analysis tasks.

For *group 2*: A *free use* for a week of VisualBib on real-world BTs, with the suggestion to consider all the main analytical and visual features of the platform; a test on *five qualitative general objectives*.

Then, both the groups were involved in a standard usability test and in providing their opinions on effectiveness of the main features and sections of the platform. Finally, only group 2 answered to a set of open-ended questions in order to collect more articulated considerations.

6.3 | Measures

In the following, we present the structure of the various parts of the user study. After each next evaluation item, we inserted in round brackets the BTs associated to them.

6.3.1 | Eight quantitative tasks

The first set of tasks was designed for group 1 to evaluate, with increasing degree of difficulty, two of the four categories of the BTs discussed in Section 3.3: *visualize and explore* and *analyze and compare*.

We proposed participants to analyze a given bibliography, asking them to execute the following eight quantitative tasks, with three levels of difficulties (easy, medium, not-trivial) related to the numbers of steps (resp. 1, 2 and 3) necessary to answer.

- T1: the number of authors in the bibliography (BT2);
- T2: the maximum number of authors per paper (BT12);
- T3: the most frequent subject area of the papers (BT6);
- T4: the author with the highest impact index (h-index) (BTs 10,12);

T5: the most frequent keyword among the first attributed to papers (BTs 6,10);

T6: the average h-index of the authors of a specific paper (BT12);

- T7: the number of authors of conference papers (BT5);

T8: the number of papers associated with two specific subject areas (BTs 6,12).

6.3.2 | Five qualitative general objectives

The second test was designed for group 2 to evaluate the effectiveness of VisualBib to pursue five general objectives in a five-point Likert-scale (1—*low effectiveness*; 5—*high effectiveness*).

O1: analysis of the scientific production of an author (BTs 1,2,4-7,10,12);

O2: discovery of specific information about the bibliography (e.g., distribution of papers over time, no. of collaborations, ...) (BTs 4-10,12);

O3: construction of a new bibliography (BTs 1,2,8,9,13);

O4: refinement of a bibliography (BTs 1-3,8,10-12);

O5: evaluation of a bibliography (BTs 2,4-7,10,12).

6.3.3 | Nine common questions for a standard usability test

Both groups were then asked to answer to standard SUS (Brooke, 2013) questionnaire, in a five-point Likert-scale, in its simplified SUS_{-01} version (without the first item) on the basis of work (Lewis & Sauro, 2017) which proofs that, when leaving out the first question, the mean difference from the score of the full SUS survey of -0.66 points, with a 95% confidence interval. We report below the questions of the SUS_{-01} :

Q2: I found the system unnecessarily complex;

Q3: I thought the system was easy to use;

Q4: I think that I would need the support of a technical person to be able to use this system;

Q5: I found the various functions in this system were well integrated;

Q6: I thought there was too much inconsistency in this system;

Q7: I would imagine that most people would learn to use this system very quickly;

Q8: I found the system very awkward to use;

Q9: I felt very confident using the system;

Q10: I needed to learn a lot of things before I could get going with this system.

6.3.4 | Seven common questions on main features

Finally, both groups were asked to provide their opinions, in a five-point Likert-scale (1—*low*; 5—*high*), about

the main features of VisualBib (E1–E4), the effectiveness of the specific sections of the user interface (E5–E6), and the overall appreciation of the system (E7):

E1: the use of narrative diagrams and related views for representing bibliographies (BTs 2,4–7);

E2: the procedure for importing and enriching a bibliography (BTs 1,8,13);

E3: the procedure for resolving authors' duplications in a bibliography (*Match Authors*) (BT9);

E4: the procedures for integrating the bibliography with further papers by an author and cited/citing papers (BTs 1,8);

E5: the use of the exploration panel (BTs 2,10,12);

E6: the analysis panel for the comparative visualization of the metrics of papers and authors (BT12);

E7: the degree of overall appreciation of the platform.

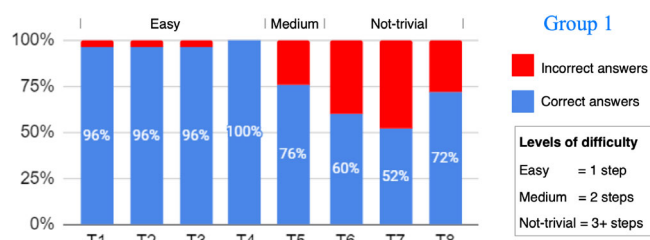


FIGURE 17 The rate of the correct answers to the tasks T1... T8 [Color figure can be viewed at wileyonlinelibrary.com]

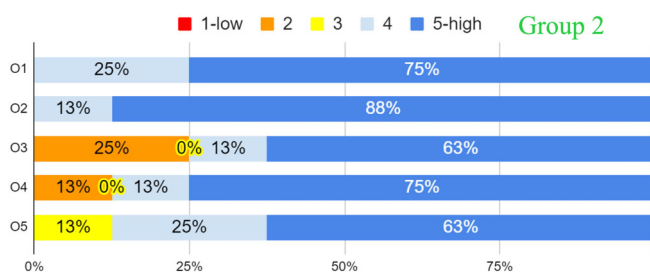


FIGURE 18 Group 2: The distribution of answers, in a scale from 1 (low effectiveness) to 5 (high effectiveness), for each of the considered objectives [Color figure can be viewed at wileyonlinelibrary.com]

6.4 | Results

For each measure presented above, we report here the results of the user study.

6.4.1 | Eight quantitative tasks

Figure 17 shows the percentage rate of correct answers to the tasks T1...T8. T1, T2, T3, and T4 received $\geq 96\%$ of correct answers; T4 involved the use of one of the filtering sliders in the keyword section of the exploration panel, and scored 76% of correct answers. The remaining tasks implied the finding of second-level metrics that required to perform at least three operations. The rates of correct answers reveal the difficulties for users to carry out nontrivial tasks without a specific training: Users participating in the study had less than an hour to experiment the platform before trying the test.

6.4.2 | Five qualitative general objectives

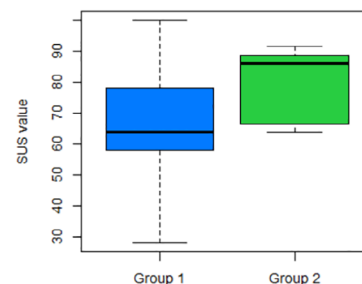
Figure 18 shows the distribution of answers about the effectiveness for general objectives. The results emphasize the strength of analytical features of the platform and its potential to uncover aspects, which are not immediately apparent in a bibliography. The support of the system in the construction of a new bibliography and its evaluation were considered slightly less than the other aims. Probably users have experienced some difficulties in creating a new bibliography starting from scratch, since the platform does not enable a free search, requires the Doi of a seed paper, and does not offer aggregated metadata in this initial phase.

6.4.3 | Nine common questions for a standard usability test

Figure 19 illustrates the comparison between the distributions of the SUS_{-01} for the two groups. The medians of the distributions of the SUS_{-01} ranges between

FIGURE 19 SUS_{-01} test results: The parameters of the distributions for the two groups [Color figure can be viewed at wileyonlinelibrary.com]

Metrics/Platforms	Group 1	Group 2
Sample size	25	8
Min	28.0	63.9
1st Qu.	58.0	66.7
Median	64.0	86.1
Mean	65.0	79.9
Std. dev.	19.1	11.9
3rd Qu.	78.0	88.9
Max	100.0	91.7



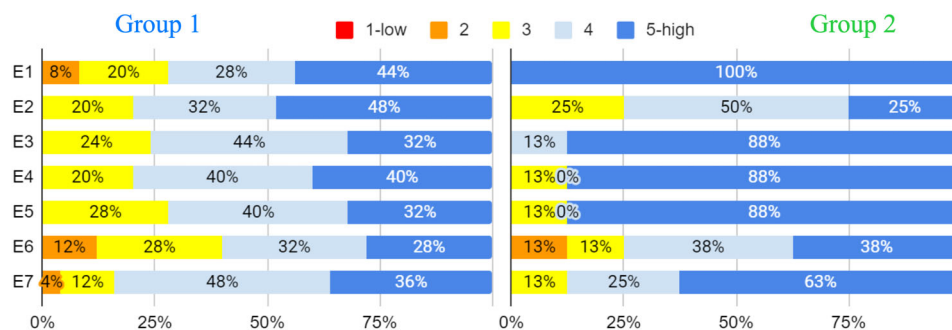


FIGURE 20 Groups 1 and 2: The distribution of perceived effectiveness of sections and procedures of the platform. E7 refers to the overall appreciation of the platform [Color figure can be viewed at wileyonlinelibrary.com]

64 (acceptable) for group 1 to 86 (excellent) for group 2 (Brooke, 2013) with a difference of 22 points: Experienced users perceived the system significantly more usable than group 1.

Analyzing more in detail, the distribution of the individual answers, for the group 1, the most positive results regarded Q6 (92% of the sample agree or strongly agree) and Q5 (84% of positive answers). The answer Q4 received the most negative responses (40% of positive feedback, although as many as 32% said to be neutral). This is probably due to the difficulty in approaching bibliographic analysis tasks, a rather uncommon task for part of the participants.

For the group 2, Q4 highlights a higher confidence with the system (75% of positive feedbacks), while Q6 got even 100%.

6.4.4 | Seven common questions on main features

About the effectiveness of sections and features, the results in Figure 20 highlights that all aspects, except E6, were positively assessed gaining at least 72% of positive scores (levels 4 or 5 of the scale) in group 1 and 75% in group 2. For both the groups, the most positive evaluations were about the overall platform (respectively 84% and 88% of positive scores); the least positive evaluation (60% in group 1) was about the radar section of the application (E6), probably not so easy to interact with for a part of the participants. The group of experts particularly appreciated the narrative diagrams (100% of higher score) and all other functions of VisualBib, including the analysis panel (E6, with 76% of appreciation).

Finally, we collected user's opinions from the expert group about strengths and weaknesses of the system, the most interesting information provided and any suggestions for new features. Among the *strengths*, six out of eight experts appreciated the views provided: keywords, subject areas, and tags, the visualization of the network of collaborations between authors and the citation network between papers. The other two experts highlighted

the intuitive use and originality of the system. *Weaknesses* that emerged included: lack of a search for papers by title or keywords on citation indexes, lack of initial guidance on functionality, limited scalability, reliance on a nonopen access database. Among the most interesting information provided they cited: the network of authors and citations, publication timeline, comparison of article and author metrics, and the distribution across the timeline of the word clouds of keywords and subject areas.

7 | CONCLUSIONS AND FUTURE WORK

In this paper, we presented VisualBib, a real-time visual analytics tool, which supports researchers in performing BTs and complex visual analyses on scientific bibliographies. The innovativeness of our proposal is in - the variety of metadata formats and of contextual visualization methods: they simplify the integration of several functions specialized in the construction and analysis of a bibliography, and in the realizations of BTs; - in providing support to researcher in the real-time retrieval of detailed DIEs and AIEs and in the immediate computing and visualization of aggregated metadata (BIEs) such as collaboration and citation networks, distributions of keywords, subject area, and tags. Furthermore, in the new proposed document use model and the new typologies metadata (AIEs and interactive, aggregated BIEs), which offer new insights in the field of literature search and review. The quantitative and qualitative user evaluations presented in this work, carried out on two different groups, received positive feedbacks, especially from the experienced, involved users, who appreciated the effectiveness of the system in supporting high-level tasks. The underlying data model, coupled with the use of advanced graphic organizers, strongly connects data along multiple dimensions, and leads to original visual representations and modalities of navigation. The proposed narrative view, with the explicit representations of contextual links, has been particularly appreciated by the researchers who tried the system.

Some interesting insights came from the expert group, who suggested to introduce a manual edit of metadata, a new advanced search on bibliographic indexes, additional BIEs, and metrics like explicit self-citation counters and venues distributions, to support the detection of papers and author clusters in a bibliography. We are already working on these improvements, and also on new features: We planned to apply automatic citation classifiers, in order to support the researchers in the semantic tagging and paper annotations; furthermore, our aim is to expand the list of the bibliographic indexes to query, and introduce new smart modalities for importing and sharing bibliographies.

ACKNOWLEDGMENT

Open Access Funding provided by Università degli Studi di Udine within the CRUI-CARE Agreement.

ORCID

Antonina Dattolo  <https://orcid.org/0000-0002-8511-524X>

Marco Corbato  <https://orcid.org/0000-0003-0723-8241>

REFERENCES

- Bergström, P., & Atkinson, D. *Augmenting the exploration of digital libraries with web-based visualizations*. Fourth International Conference on Digital Information Management, 1–7, 2009. <https://doi.org/10.1109/ICDIM.2009.5356798>.
- Bethard, S., & Jurafsky, D. *Who should i cite? Learning literature search models from citation behavior*. Proceedings of the 19th ACM International Conference on Information and Knowledge Management – CIKM'10, 609–617, 2010. <https://doi.org/10.1145/1871437.1871517>.
- Bikakis, N., Maroulis, S., Papastefanatos, G., & Vassiliadis, P. (2021). In-situ visual exploration over big raw data. *Information Systems*, 95, 101616. <https://doi.org/10.1016/j.is.2020.101616>
- Booth, A., Sutton, A., & Papaioannou, D. (2016). *Systematic approaches to a successful literature review* (second ed.). Sage ISBN 9780857021359.
- Borgman, C., & Furner, J. (2002). Scholarly communication and bibliometrics. *Annual Review of Information Science and Technology*, 36, 2–72.
- Bostock, M., Ogievetsky, V., & Heer, J. (2011). D³ data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, 17(12), 2301–2309.
- Bourhis, P., Reutter, J. L., Suárez, F. & Vrgoč, D. (2017). Proceedings of the 36th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, pp. 123–135, <https://doi.org/10.1145/3034786.3056120>
- Brooke, J. (2013). Sus: A retrospective. *Journal of Usability Studies*, 8(2), 29–40.
- Chen, C., Ibekwe-SanJuan, F., & Hou, J. (2010). The structure and dynamics of co-citation clusters: A multiple-perspective co-citation analysis. *Journal of the American Society for Information Science and Technology*, 61(7), 1386–1409. <https://doi.org/10.1002/asi.21309>
- Chen, J., & Zhuge, H. (2019). Automatic generation of related work through summarizing citations. *Concurrency Computation*, 31(3), 12. <https://doi.org/10.1002/cpe.4261>
- Costagliola, G., De Rosa, M., Fuccella, V., & Perna, S. (2018). Visual languages: A graphical review. *Information Visualization*, 17(4), 335–350. <https://doi.org/10.1177/1473871617714520>
- Costagliola, G., & Fuccella, V., *Cybis: A novel interface for searching scientific documents*. 15th International Conference on Information Visualisation, pages 276–281, 2011. ISBN 9781457708688. <https://doi.org/10.1109/IV.2011.95>.
- Dattolo, A., & Corbato, M. (2019). Visualbib: A novel web app for supporting researchers in the creation, visualization and sharing of bibliographies. *Knowledge-Based Systems*, 182, 104860. <https://doi.org/10.1016/j.knosys.2019.07.031>
- Dattolo, A., & Luccio, F. L. *A formal description of zz-structures*. 1st workshop on new forms of Xanalogical storage and function, CEUR(508), pages 7–11, 2009a.
- Dattolo, A., & Luccio, F. L. *A state of art survey on zz-structures*. 1st workshop on new forms of Xanalogical storage and function, CEUR(508), pages 1–6, 2009b.
- Di Iorio, A., Giannella, R., Poggi, F., & Vitali, F. *Exploring bibliographies for research-related tasks*. 24th International Conference on World Wide Web, pages 1001–1006, 2015. <https://doi.org/10.1145/2740908.2742018>.
- Dong, A., Zeng, W., Chen, X., & Z. Cheng. Vistory: Interactive storyboard for exploring visual information in scientific publications. In Proc. of the 12th Int. Symposium on Visual Information Communication and Interaction, VINCI'2019. ACM, 2019. ISBN 9781450376266. <https://dl.acm.org/doi/10.1145/3356422.3356430>
- Dork, M., Riche, N. H., Ramos, G., & Dumais, S. (2012). Pivotpaths: Strolling through faceted information spaces. *IEEE Transactions on Visualization and Computer Graphics*, 18(12), 2709–2718. <https://doi.org/10.1109/TVCG.2012.252>
- Elmqvist, N., & Tsigas, P. (2007). Citewiz: A tool for the visualization of scientific citation networks. *Information Visualization*, 6(3), 215–232. <https://doi.org/10.1057/palgrave.ivs.9500156>
- Federico, P., Heimerl, F., Koch, S., & Miksch, S. (2017). ISSN 1077-2626). A survey on visual approaches for analyzing scientific literature and patents. *IEEE Transactions on Visualization and Computer Graphics*, 23(9), 2179–2198. <https://doi.org/10.1109/TVCG.2016.2610422>
- He, J., Ping, Q., Lou, W., & Chen, C. (2019). Paperpoles: Facilitating adaptive visual exploration of scientific publications by citation links. *Journal of the Association for Information Science and Technology*, 70(8), 843–857. <https://doi.org/10.1002/asi.24171>
- Jeong, C., Jang, S., Park, E., & Choi, S. (2020). A context-aware citation recommendation model with bert and graph convolutional networks. *Scientometrics*, 124(3), 1907–1922. <https://doi.org/10.1007/s11192-020-03561-y>
- Jurgens, D., Kumar, S., Hoover, R., McFarland, D., & Jurafsky, D. (2018). Measuring the evolution of a scientific field through citation frames. *Transactions of the Association for Computational Linguistics*, 6, 391–406.
- Kim, J., & Kim, J. (2020). Effect of forename string on author name disambiguation. *Journal of the Association for Information Science and Technology*, 71(7), 839–855. <https://doi.org/10.1002/asi.24298>
- Kucher, K., & Kerren, A. Text visualization techniques: Taxonomy, visual survey, and community insights. *IEEE Pacific Visualization Symposium*, 117–121, 2015. <https://doi.org/10.1109/PACIFICVIS.2015.7156366>.
- Lee, B., Czerwinski, M., Robertson, G., & Bederson, B. B. *Understanding research trends in conferences using paperlens*. In CHI

- 2005 EA, pages 1969–1972. ACM, 2005. ISBN 1-59593-002-7. <https://doi.org/10.1145/1056808.1057069>.
- Lewis, J. R., & Sauro, J. (2017). Can I leave this one out? The effect of dropping an item from the SUS. *Journal of Usability Studies*, 13(1), 38–46.
- Matejka, J., Grossman, T., & Fitzmaurice, G. (2012). ISBN 978-1-4503-1016-1). Citeology: Visualizing paper genealogy. In *CHI '12 extended abstracts on human factors in computing systems, CHI EA '12* (pp. 181–190). ACM. <https://doi.org/10.1145/2212776.2212796>
- McGuffin, M., & Schraefel, M. C. A comparison of hyperstructures: Zstructures, mspaces, and polyarchies. *Proceedings of the ACM Conference on Hypertext*, pages 153–162, 2004.
- Nelson, T. H. (2004). A cosmology for a different computer universe: Data model, mechanisms, virtual machine and visualization infrastructure. *Journal of Digital Information*, 5(1) ISSN 1368-7506.
- Ponsard, A., Escalona, F., & Munzner, T. Paperquest: A visualization tool to support literature review. *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA '16*, 2264–2271, 2016. ACM. ISBN 9781450340823. <https://doi.org/10.1145/2851581.2892334>.
- Shen, Z., Ogawa, M., Teoh, S. T., & Ma, K.-L. Biblioviz: A system for visualizing bibliography information. *Asia-Pacific Symposium on Information Visualisation 60*, pages 93–102, 2006. Australian Computer Society ISBN 1-920682-41-4.
- Sultanum, N., Murad, C., & Wigdor, D. *Understanding and supporting academic literature review workflows with LitSense. ACM International Conference Proceeding Series*, 2020. <https://doi.org/10.1145/3399715.3399830>.
- Tenopir, C., Wang, P., Zhang, Y., Simmons, B., & Pollard, R. (2008). Academic users' interactions with sciencedirect in search tasks: Affective and cognitive behaviors. *Information Processing and Management*, 44(1), 105–121. <https://doi.org/10.1016/j.ipm.2006.10.007>
- Van Eck, N., & Waltman, L. (2010). Software survey: Vosviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538. <https://doi.org/10.1007/s11192-009-0146-3>
- Van Eck, N. J., & Waltman, L. (2014). Citnetexplorer: A new software tool for analyzing and visualizing citation networks. *Journal of Informetrics*, 8(4), 802–823. <https://doi.org/10.1016/j.joi.2014.07.006>
- van Eck, N. J., & Waltman, L. (2017). Citation-based clustering of publications using citnetexplorer and vosviewer. *Scientometrics*, 111(2), 1053–1070. <https://doi.org/10.1007/s11192-017-2300-7>
- Vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., & Cleven, A. *Proceedings of the 17th European Conference on Information Systems (ECIS). Verona, Italy, Jun 8-10, 2009*. <https://aisel.aisnet.org/ecis2009/161>.
- Wang, P., & Soergel, D. (1998). A cognitive model of document use during a research project. Study i. document selection. *Journal of the American Society for Information Science*, 49, 115–133.
- Zeng, W., Dong, A., Chen, X., & Cheng, Z.-L. (2020). Vistory: Interactive storyboard for exploring visual information in scientific publications. *Journal of Visualization*, 24, 69–84. <https://doi.org/10.1007/s12650-020-00688-1>
- Zhang, G., Wang, J., Liu, J., & Pan, Y. (2021). Relationship between the metadata and relevance criteria of scientific data. *Data Science Journal*, 20(1), 1–21. <https://doi.org/10.5334/dsj-2021-005>
- Zhang, X., Qu, Y., Giles, C., & Song, P. (2008). *CHI '08: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Florence Italy, April 5 - 10* (pp. 677–680). <https://doi.org/10.1145/1357054.1357161>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Dattolo, A., & Corbato, M. (2021). Assisting researchers in bibliographic tasks: A new usable, real-time tool for analyzing bibliographies. *Journal of the Association for Information Science and Technology*, 1–20. <https://doi.org/10.1002/asi.24578>