

Supporting Historical Research Through User-Centered Visual Analytics

Nadia Boukhelifa, Emmanouil Giannidakis, Evanthia Dimara, Wesley Willett and Jean-Daniel Fekete

INRIA Saclay

Abstract

In this paper we describe the development and evaluation of a visual analytics tool to support historical research. Historians continuously gather data related to their scholarly research from archival visits and background search. Organising and making sense of all this data can be challenging as many historians continue to rely on analog or basic digital tools. We built an integrated note-taking environment for historians which unifies a set of functionalities we identified as important for historical research including editing, tagging, searching, sharing and visualization. Our approach was to involve users from the initial stage of brainstorming and requirement analysis through to design, implementation and evaluation. We report on the process and results of our work, and conclude by reflecting on our own experience in conducting user-centered visual analytics design for digital humanities.

Categories and Subject Descriptors (according to ACM CCS): H.5.2 [Information Interfaces and Presentation]: User Interfaces—User-centered design

1. Introduction

Many historians choose to visit archives and libraries to conduct innovative primary research. This close proximity to sources, both physical and intellectual, allows them to quickly identify the range of material within collections, to access relevant documents, and to interpret their contents productively [Cun15]. During these archival visits, historians consult written word such as diaries, letters and notebooks, as well as film, photographs and sound collections. They take notes and photos, and transcribe resources relevant to their research questions. As archive opening hours are limited, and travelling to remote places incurs costs, historians aim to acquire as much information and resources as possible in-situ. Later, they organise, clean and analyse their data to finally produce a report of their findings leading to an eventual publication. This process of *inquiry*, *analysis* and *synthesis* can be facilitated by digital and collaborative tools, yet many historians continue to work solo and rely on analog paper or simple text processing [GO12].

The aim of this work is to build an online collaborative platform to support archival research. To this end, we built an integrated Note-Taking Environment (NTE) for historians which offers facilities for organising, authoring, sharing

and sense-making of resources related to historical research. The design and implementation of this tool was largely influenced by our close collaboration with a large number of historians working in the same project and outside. We followed a user-centered design approach where we continuously met and exchanged with a large number of historians and related scholars.

Our contributions are three-fold: (1) a detailed analysis of user requirements through three participatory design sessions, (2) an implementation of a prototype Note-Taking Environment (NTE), and (3) two user evaluations of the NTE. This work is part of a larger European project (CENDARI [CEN15]) that brings information and computer scientists together with leading historians and existing historical research infrastructures to improve the conditions for historical scholarship in Europe.

2. Understanding Historical Research Through Participatory Design

Although our focus is to facilitate early research for historians, other user groups are also implicated in archival research such as librarians and archivists. We organised three participatory design sessions, inspired from [BLM03], with

three different user groups (Fig. 1): WWI historians, librarians and medievalists. We expected these to have different interests, practices and goals of research. The aim of these sessions was to understand how different user groups would want to search, browse, and visualize (if at all) information from archival research. In total we had 49 participants (14 in the first, 15 in the second and 20 in the third).



Figure 1: *The three participatory design sessions held.*

2.1. Methodology

Each session run as a one-day workshop and was divided into two parts. The morning began with presentations of existing interfaces for access and visualisation. In order to brainstorm productively in the afternoon, participants needed to have a clear idea of the technical possibilities currently available. In the afternoon, participants divided into 3 – 5 groups of four and brainstormed ideas for searching, browsing, and visualization functions, and then they created paper and video prototypes for their top three ideas. Everyone then met to watch and discuss the videos.

2.2. Results

Even though there were different groups of users involved in each workshop, these common three themes emerged from the discussions and the prototypes:

T1: Note-taking as a central activity: this could be paper-based or in a digital form with certain types of note-taking already heavily computer-based (e.g. transcription of financial records).

T2: Privacy as a main concern: participants expressed reluctance to share their notes, since they are highly personal and tailored to the current topic of research. They are also very interpretive as one researcher noted: “*History is not about information, it is about sources and the choices you make as a historian*”. Notes reflect a historian’s interpretation of the source and can therefore be problematic in terms of usefulness to other researchers.

T3: Entities as the basic unit of information sharing: participants were interested in sharing certain types of information such as named entities (e.g. for person, organisation) and skeptical about the value of sharing other types (e.g. notes). Sharing is more likely in an already existing network (e.g. of friends and colleagues), unless there is an important or altruistic cause such as reconstructing a lost archive.

T4: Search as a ubiquitous tool: participants wanted to

be able to search for different types of documents (e.g. notes, documents, transcripts, scans etc) and find entities within these documents. They also wanted to search by different themes such as by language, period or concept.

T5: Visualization as a useful asset: from the video prototypes and the discussion, participants were very interested in visualizations that would allow them to conceptualise information in ways that are difficult in text-based forms.

2.3. User Requirements

Based on the participatory design sessions and discussing with historians, we delineated five main user requirements:

- [R1] Support editing of notes and documents
- [R2] Allow user control over privacy and sharing
- [R3] Allow tagging of text to create named entities
- [R4] Provide a robust search facility
- [R5] Provide visualization tools

3. Related Work

Work that relates to ours can be summarised in two parts: note-taking tools and entity-based visualizations.

Note-taking Environments: we focus here on digital note taking tools as opposed to paper-based ones. Indeed, Walsh et al. [WC12] found that that electronic laboratory notebooks outperform their paper-based counterparts in an academic research setting. There is a myriad of digital note-taking tools relevant to historians. Some of these tools focus on citation management (e.g. EndNote [Reu15], Zotero [Zot15]) and collection and organization of notes (e.g. Evernote [Eve15]). Others focus on integration of analog and handwritten notes (e.g. LiveScribe [Liv15]), annotation (e.g. Reader’s Notebook [FXP15]), collaboration (e.g. Google-Docs), search (e.g. CoSense [PM09]) and/or visualization (e.g. CommentSpace [WHHA11]). Tools that are particularly built for historical research include Scribe [Scr15] and Editors’ Notes [Not15]. Scribe is targeted for historians and uses digital note cards for research notes, quotes, sources, digital images and outlines. Editors’ Notes, on the other hand, is designed for editors, archivists and librarians. It is a web-based tool for recording, organising and preserving research notes. Although Editors’ Notes has a rich set of functions accessible via an API, they do not currently offer important features such as visualization, faceted search or rich semantics for research notes.

Entity-based Visualization: similar to our tool, Jigsaw [SGL08] works with collections of text documents, uses multiple coordinated views and emphasises connections between entities across documents. D-Dupe [BLGS05] is another entity-based visualization system designed to help resolve duplicated entities in social networks. Similar to our work, through highlighting and linking of entities across documents, we hope to enable users to identify ambiguities and conflicts, and to resolve them. Finally, FacetAtlas

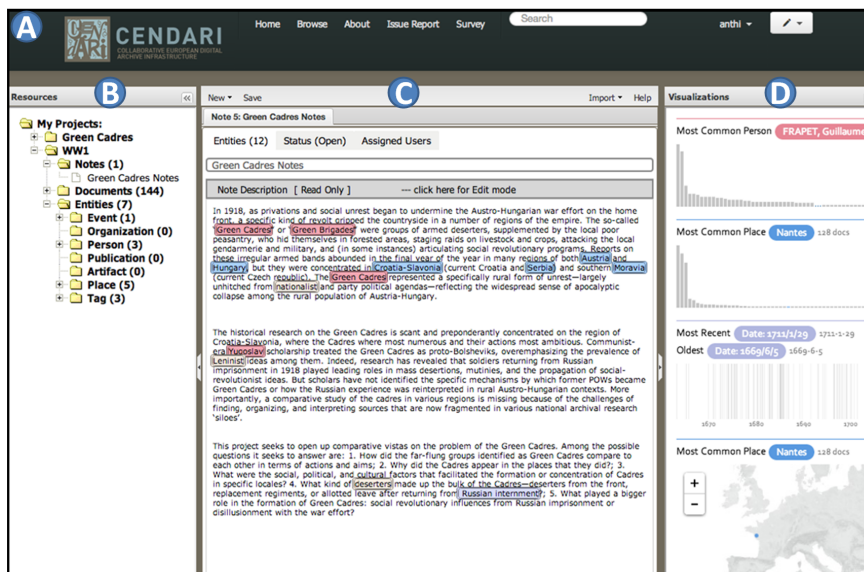


Figure 2: The Integrated Note-taking Environment NTE for Historians: (A) the search and browse panel, (B) the resources panel, (C) the central panel dedicated to editing, tagging and resolution, and (D) the visualization panel.

[CSL*10] focuses on multi-faceted relationships between documents.

Unlike general note-taking tools, our aim is to provide a rich integrated environment tailored for historians. The environment we envisage includes all of the important features listed earlier, such as note-taking, annotation, search, visualization and collaboration. Our approach is to build on existing tools and to take a user-centered design approach. We chose Editors' Notes [Not15] platform to build from as it is open-source, feature-rich and has already been adopted in various large projects for historical research.

4. An Integrated Note-Taking Environment for Historians

The goal of our Note-Taking Environment (NTE) is to support the research process of historians, allowing them to collect notes, augment them and collaborate. This section describes the design and implementation of the NTE.

4.1. Design of the NTE

The main user interface of the note taking environment has three main panels besides the search and browse panel (Fig. 2(A)); a library where the user can manage projects and browse allocated resources (Fig. 2(B)); a central space for editing, linking and tagging resources (Fig. 2(C)); and a visualization space for showing trends and relationships in the data (Fig. 2(D)).

The resources panel: resources are organised per project into three main folders that correspond roughly to the way

historians organise their material on their machines. The notes folder contains files, each file is a note describing archival material related to a project. The user can select a note and its content is shown in the central panel. The user can edit the note, tag words to create entities such as event, organisation, person, place and tag. Users can add references to documents. The documents themselves can be letters, newspaper articles, contracts or any text that acts as evidence for an observation or a statement in a note. Documents can contain named entities, a transcript, references to other documents and resources as well as scanned images, which are displayed in the high resolution image viewer at the bottom of the central panel.

The central panel acts as a viewing space for any type of resource and thus mimics the function of a working desk of a historian. This is also where entity tagging and resolution takes place. The user may not be entirely clear about the true identity of an entity, for instance, in the case of a city name that exists in different countries. The user has the option to manually resolve an entity, when further information becomes available, by assigning a unique resource identifier URI (e.g. to a unique Wikipedia entry).

The visualization panel provides useful charts to show an overview of entities, distributions, frequencies and outliers in the resources of the project. There is also a map which shows the location of any *place* entity.

The three panels are coordinated using brushing and linking. In terms of privacy settings, by default notes are private and entities are public but users can change these permissions.

In summary, our NTE supports the following functionalities: (1) **Editing and annotation**: through a rich set of editing, formatting and tagging options provided by RDFaCE [RDF15] [supporting R1,R3]; (2) **Faceted Search**: a faceted search service for thematic access to resources [Ela15] [supporting R4]; (3) **Visualization**: showing histograms of three entity types (names, places and events) as well as a geographical map with support for aggregation [supporting R5]. Other visualizations could be easily integrated; (4) **Automatic Analysis**: in the form of an automatic entity recognition integrated service [RDF15] (5) **Interaction**: the visualizations support selection, highlight and pan & zoom for the map. Brushing and linking is implemented with one flow direction for consistency from left to right. This is to support the workflow in the NTE: select resource, view & update, then tag and visualize. Note that referencing, collaboration and privacy setting [related to R2] are available in the NTE but we do not discuss them here as they are not part of our user evaluations.

4.2. A Note on the Implementation

We use a client-server architecture; on the client side we rely on modern web technologies (HTML5, JavaScript, D3.js) and Django web framework on the server side. Extracted entities are turned into RDF triples organized through a *CEN-DARI*-specific ontology. The NTE faceted browsing and search functionalities are implemented using Elasticsearch [Ela15], which unifies the exploration of resources provided by the project (text, HTML, XML).

5. User Evaluation

We conducted two user evaluations in order to: (1) evaluate the basic design of the NTE interface, (2) assess participants performance in carrying out the main tasks (upload, organise, tag, resolve and use the visualizations), and (3) gather user feedback for the next iteration. Each study was organised in two parts, a short introduction to the tool, then a hands-on part where participants worked on their own data. Prior to each workshop we asked participants to bring material related to a research project (e.g. personal notes, documents and scans). During each session, two study facilitators were present to answer questions. Each session lasted around two hours, after which participants were asked to fill in an online survey where they ranked features of the NTE in terms of *usefulness* and *ease of use*. We had 18 participants in the first workshop and 14 in the second one. Participants were young historians, mostly PhD students interested in using digital tools to conduct their research. Their experience with digital humanities tools varied from novice to some users indicating basic familiarity with certain tools.

Results from the first study showed that participants were very enthusiastic about using the tool and ideas behind it. They commented positively about the intuitive layout and the ability to resize the different areas (e.g. when writing

notes, participants preferred to maximise the central space). Importantly, they quickly grasped the underlying model for organising resources into projects, notes, documents and topics. Surprisingly, participants were eager to tag entities as they saw the immediate reward when the visualizations started to show data and trends. There was also a positive attitude towards sharing. However, participants complained about the difficulty in navigating between the different types of resources (notes, documents and topics). This was partly due to technical issues concerning the response time of the system and the update mechanism between the different panels. We tried to address some of these issues for the second study which took place six months later. The results of the questionnaire from the second study showed that participants appreciated the feature-rich aspect of our tool, in particular tagging & resolution, linking of resources through entities and the visualizations. Overall participants thought that the user interface is intuitive, somewhat easy to use and that the NTE is useful to the way historians conduct their research (Fig. 3).

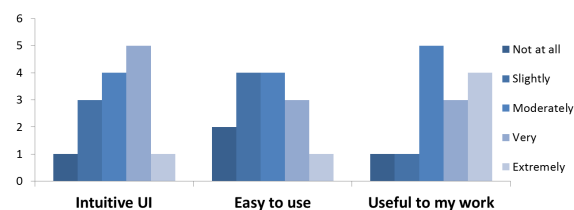


Figure 3: Survey results of study 2.

6. Concluding Remarks

We followed a user-centered approach to design and implement a note-taking environment for historians. Our long-term goal is to bring value to research notes, foster collaboration and sharing of documents between historians. To achieve this, we had to understand how historians work and identify suitable opportunities for sharing and collaboration. The idea of *selective sharing* [BOD14] via tagging and linking of entities seems to answer to a basic need for awareness and collaboration. It is also a step forward towards building a community-cleaned repository of historical knowledge.

Our experience working closely with historians showed us the importance of involving users early in the project, to be able to understand their needs and identify key challenges and requirements. The domain of digital humanities offers exciting opportunities for visual analytics. However, before any digital advances are to be embraced by this community, technological and psychological barriers need to be overcome. The lack of adoption has remained a challenge partly because of the limited accessibility of our tools which can be attributed to the paucity of appropriate documentation and training material [GO12]. Our note-taking environment is available on Github [NTE15] and we are working on creating support documentation and video tutorials to improve our tool adoption in the digital humanities community.

References

- [BLGS05] BILGIC M., LICAMELE L., GETOOR L., SHNEIDERMAN B.: D-dupe: An interactive tool for entity resolution in social networks. In *International Symposium on Graph Drawing* (September 2005), Healy P., Nikolov N. S., (Eds.), vol. 3843 of *Lecture Notes in Computer Science*, Springer, pp. 505–507. [2](#)
- [BLM03] BEAUDOUIN-LAFON M., MACKAY W.: Prototyping tools and techniques. In *The Human-computer Interaction Handbook*, Jacko J. A., Sears A., (Eds.). L. Erlbaum Associates Inc., Hillsdale, NJ, USA, 2003, pp. 1006–1031. [1](#)
- [BOD14] BENEŠ J., O’CONNOR A., DIMARA E.: The CENDARI Project: A user-centered enquiry environment for modern and medieval historians [Poster]. In *Digital Humanities* (Lausanne, Switzerland, July 2014), pp. 434–436. [4](#)
- [CEN15] CENDARI: Collaborative european digital archive infrastructure, 2015. [1](#)
- [CSL*10] CAO N., SUN J., LIN Y.-R., GOTZ D., LIU S., QU H.: FacetAtlas: Multifaceted visualization for rich text corpora. *Visualization and Computer Graphics, IEEE Transactions on*, 16, 6 (Nov 2010), 1172–1181. [3](#)
- [Cun15] CUNNINGHAM S.: Archive skills and tools for historians, 2015. [1](#)
- [Ela15] ELASTICSEARCH: Elasticsearch, 2015. [4](#)
- [Eve15] EVERNOTE: Evernote, 2015. [2](#)
- [FXP15] FXPAL: Reader’s notebook, 2015. [2](#)
- [GO12] GIBBS F., OWENS T.: Building better digital humanities tools: Toward broader audiences and user-centered designs. *Digital Humanities Quarterly* 6 (2012). [1](#), [4](#)
- [Liv15] LIVESCRIBE: Livescribe, 2015. [2](#)
- [Not15] NOTES E.: Editors’ notes, 2015. [2](#), [3](#)
- [NTE15] NTE C.: The Note-Taking Environment NTE Repository, 2015. [4](#)
- [PM09] PAUL S., MORRIS M. R.: Cosense: Enhancing sense-making for collaborative web search. [2](#)
- [RDF15] RDFACE: Rdface, 2015. [4](#)
- [Reu15] REUTERS T.: Endnote, 2015. [2](#)
- [Scr15] SCRIBE: Scribe, 2015. [2](#)
- [SGL08] STASKO J., GÖRG C., LIU Z.: Jigsaw: Supporting investigative analysis through interactive visualization. *Information Visualization* 7, 2 (Apr. 2008), 118–132. [2](#)
- [WC12] WALSH E., CHO I.: Using evernote as an electronic lab notebook in a translational science laboratory. *Journal of Laboratory Automation* (Dec. 2012). [2](#)
- [WHHA11] WILLET W., HEER J., HELLERSTEIN J., AGRAWALA M.: Commentspace: Structured support for collaborative visual analysis. *ACM Human Factors in Computing Systems (CHI)* (2011), 3131–3140. [2](#)
- [Zot15] ZOTERO: Zotero, 2015. [2](#)