Rainer Simon*, Joachim Korb**, Christian Sadilek***, Matthias Baldauf****

Explorative user interfaces for browsing historical maps on the Web

Keywords: Historical GIS; collaboration; crowdsourcing; Tag Clouds.

Summary:
Libraries and cultural heritage institutions are increasingly making efforts to digitise their holdings for the purposes of archiving and online publication. Historical maps are a particularly interesting category of cultural heritage material: They are not only an illustration of accumulated geographical knowledge of the time; they also draw a fascinating picture of the cultural, political, scientific, religious and mythological context in which they were created. To the trained scholar, characteristics such as the style of cartographic representation, text inscriptions and legends, symbols and ornamental features, or specific geographical errors and misconceptions reveal a wealth of historical background information. To the student or the layperson, however, this information is not easily accessible.
In this paper we propose a Web-based system that will make it possible for scholars to collaboratively work together in the interpretation of these features. To the general user, this information will become easily accessible when interpretations are made public. Thus the system represents a collaborative academic tool, as well as a showcase of academic work. Community involvement, joint aggregation and filtering of user contributed as well as externally linked data, and a user interface that invites exploration constitute the cornerstones of our envisioned system: It enables users to add annotations to the map as a whole or parts of it; it provides basic GIS features such as (approximate) collaborative geo-referencing, feature search, import/overlay and export of point and line data; and it visualises aggregated context information in the form of tag clouds which emerge around the user's mouse pointer. We describe goals and functionalities we envision for the proposed system, present early proof-of-concept implementations, and discuss areas of future work.

Introduction

Old maps are a fascinating part of our cultural heritage. On the one hand, they document the state of geographical and scientific knowledge of their time. On the other hand, they provide insight into the cultural, political, religious and mythological context in which they were created. Historical maps can rightfully be considered as part of the artistic heritage as much as they are part of the history of science and technology in general (Boutoura and Livieratos 2006).
In this paper, we report on work in progress carried out as part of the EU-funded *EuropeanaConnect* project. One of the goals of this project is the creation of portal technologies that enable community involvement on the *Europeana* digital cultural heritage Web portal. Our specific interest in the project is the design of services and user interfaces that enable public participation by means of media annotation: Annotations provide valuable meta-data which can improve search and retrieval on the portal, and aid others with the interpretation and understanding of a particular cultural heritage artifact. Beyond that, however, we argue that they can also serve as a general

* Digital Memory Engineering, Austrian Institute of Technology [rainer.simon@ait.ac.at]
** Digital Memory Engineering, Austrian Institute of Technology [joachim.korb@ait.ac.at]
*** Digital Memory Engineering, Austrian Institute of Technology [christian.sadilek@ait.ac.at]
**** Telecommunications Research Center Vienna [baldauf@ftw.at]

medium for knowledge exchange and communication among users of the portal; they can facilitate community formation; and ultimately lead users to engage with the matter more deeply.

It is this fundamental assumption about the value of annotations which also forms the background of our work with historical maps: We are interested in exploring how collaboration and community involvement can help in facilitating a better and more comprehensive understanding of maps as cultural heritage artifacts.

In the following sections, we introduce our vision of a Web-based tool for the collaborative analysis, interpretation and exploration of digitised historical maps. We propose four essential functionalities as the tool's conceptual basis: *annotation*, *geo-referencing*, *synthesis* and *exploration*. We discuss possible features and pre-requisites needed to support them in a practical setting; and we present our ongoing efforts with regard to implementing software prototypes to demonstrate our vision. We conclude the paper with an outlook on future work and upcoming research activities.

A Collaborative Environment for Analysing & Experiencing Historical Maps

To the trained scholar and map expert, characteristics such as the style of cartographic representation, text inscriptions, legends, decorative elements, or specific geographical errors, misconceptions or myths depicted on the map reveal a wealth of historical context information. To members of the general public with no prior knowledge or domain-specific education, however, this information remains inaccessible.

In our vision, the role of an online cultural heritage portal is exactly to bridge this knowledge gap between scholars and the public: Visitors to the portal should be able to discover and learn about the map, as if guided by an expert. They should be pointed to particular aspects of the map which are typical for the time, and be made aware of peculiarities which set it apart from other works. Likewise, they should be provided with a sense of the "big picture" – the larger historical context in which the map was created: What political, societal, cultural, religious situations predominated? What was the level of scientific development and geographical knowledge about other parts of the world? Which major milestone events or conflicts happened at that time and may have been influential?

Accordingly, the collaborative system for the study and exploration of historical maps that we envision acts, above all, as a platform for knowledge exchange: map experts and historians must have an incentive to contribute information; the system collects and aggregates this expert knowledge; and provides appropriate means to convey that knowledge to the visitor. Moreover, we envisage this process to be iterative, and to provide facilities for feedback and communication: Eventually, we hope for an emerging dynamic where communities may form, and the roles of "the expert" and "the visitor" begin to blur as visitors become more knowledgeable, and experts may benefit from qualified input provided by the general public.

In order to approach this broad vision in a pragmatic way and within the scope of our project, we introduce four basic functionalities which we believe such a system must provide: (1) *Annotation* – an essential functionality that allows expert as well as non-expert users to contribute knowledge to the system in a simple and effortless way. (2) *Geo-referencing* – a supporting functionality that not only allows studying some of the maps' technological aspects, but also enables a range of potentially interesting interaction features within the portal. (3) *Synthesis* – functionality to carry out comparative analyses on features of the map. (4) *Exploration* – the "functionality" which combines all the expert knowledge in the system and unlocks it for the general public by making it available under a single user interface which invites users to investigate, explore and interact. In the following sub-sections, we discuss each of the four use cases in more detail.

Annotation

Annotations are a powerful means of collaboration, both among expert and end-user communities. They make implicit information explicit, and convey additional meaning about the content to which they refer. They supplement published information, and can provide corrections that make it more appropriate to the user's setting (Frisse 1987). In the Cultural Heritage sector, annotations can serve for enriching digital resources with valuable metadata, and thus improve search and discoverability within the digital library system or on the World Wide Web (Haslhofer et al. 2009).

Within *EuropeanaConnect*, as well as in two earlier related projects[1,2], we have been investigating how media annotation (images, hypertext, video and audio) can add value to cultural heritage portals; and how a critical mass of user contributions may be reached by providing appealing, feature-rich collaboration user interfaces to cultural heritage content. The tools for annotating historical maps which we propose in this paper evolved directly out of this work, as a specialization of our image annotation toolset in particular. The experience gathered in these projects provided input to many of the envisioned system's features.

Obviously, one of the most essential features is support for annotating map fragments: Annotations may pertain to the map as a whole, or only to parts of it. Such a system would have to provide a set of tools for drawing draggable and resizable shapes, with configurable color and stroke size onto the map: points to mark locations of interest; lines to trace linear features such as rivers or routes depicted on the map; rectangles, ellipses and polygons for marking arbitrary regions, and a freehand drawing tool.

For the management of the annotations, a simple user interface component that consists primarily of a scrollable list should be provided. To avoid clutter and occlusion on the map, the annotation shapes will not be shown by default. Only if the user hovers over an annotation in the list, the corresponding shape(s) will be superimposed over the map. In case the map is geo-referenced (see following sub-section), users will be able to export annotations to an open format such as KML[3] for visualization in 3rd party mapping toolkits or virtual globe browsers.

The actual free-form annotation text, as well as "tags" (freely chosen descriptive keywords which users can assign) will also be edited in the annotation management list: The idea is to allow users to choose tags freely, or select them from a pre-defined (but extensible) controlled vocabulary. The controlled vocabulary will be particularly helpful, as it provides a simple means to define and reference categories of elements that are commonly found across maps – e.g. layout elements such as the decorative border, the actual map area, legends, or illustrative imagery.

An annotation tool should motivate discussion. Therefore, it must support reply threading to facilitate communication and information exchange between contributors. As sharing and cooperation is not always the intention of the creator of an annotation, the user should be able to define an annotation's scope to limit its visibility.

In an open, collaborative environment, mechanisms for ensuring the quality of user-contributed content are vital: Since annotations can be left in numerous languages, and the potential number of annotations created each day might be large, peer review mechanisms probably represent the

[1] BRICKS, http://cordis.europa.eu/ist/digicult/bricks.htm
[2] TELplus, http://www.theeuropeanlibrary.org/telplus
[3] http://www.opengeospatial.org/standards/kml/

only practical approach in this regard: On the one hand, users should have the possibility to indicate inappropriate annotations to the system administrator in order to report spam, vandalism or otherwise offensive or abusive behavior on the system. On the other hand, there should also be mechanisms that reward positive contributions: For example, public recognition, public ranking tables or user profiles have been reported by users as additional motivating factors under comparable conditions (Holley 2009).

Geo-Registration

When analysing historical maps, geo-registration – i.e. knowledge of a defined correspondence between the map's coordinates and a well-defined geographical coordinate system – may not necessarily be a requirement. In fact, establishing such a correspondence may be problematic, even impossible, in some cases (e.g. for some medieval mappae mundi, which are often a depiction of the religious and cultural view of the world rather than a depiction of geography).
Nevertheless, a toolset that allows for comparing and matching the geometry of a historical map with modern counterparts is useful: Not only does it provide insight into the accuracy of the map; it may also help understand its method of construction (Tobler 1966), give hints about possible underlying projections (Boutoura 2006), and can generally aid in supporting or refuting hypotheses about technical aspects of the map's creation (e.g. surveying methods or source material used (Jenny et al. 2007).
Establishing a correspondence between the image-coordinate space of a historical map with unknown projective properties and a well-known geographical projection system requires knowledge about the map: At least, a set of properly distributed *control points* – recognizable points on the historical map to which the geographical coordinates are known – must be available. These control points can then be used as a basis for translations between both coordinate spaces: e.g. by applying transformations well-documented in the geometric and geodetic literature such as the *affine*, the *projective*, the *polynomial* or the *finite element transformation* (Boutoura and Livieratos 2006). Alternatively, "trial and error" can be used to test the historical map against a known map projection model. In this case, the control points can be used to quantify the amount of agreement with the tested projection. For example, Boutoura (2006) successfully applied this approach to late medieval and early and mid-renaissance portolan maps. In her work, she points out that a key challenge with this approach is the selection of a proper projection test model, and the proper spatial distribution of control points. She also stresses the need for an interactive visualisation tool to aid the trial and error process.
To summarise, the system we envision should support the process of geo-registration by providing the ability to create and edit control points; and by offering computational facilities to perform translations between the map's image-coordinate space and geographical reference systems. Our system should support experimentation with different projection models through a suitable user interface for selecting and parameterising projections, and through appropriate visual feedback, e.g. in the form of superimposed coordinate graticule or country outlines. It is worth pointing out that the main benefit of the envisioned system, compared to existing tools, is collaboration: Through collaboration, the problems pointed out by Boutoura may potentially be overcome. The community could be involved in gathering control point data, or in providing controlling input to the trial and error process of searching for appropriate map projections. This approach may seem problematic at first, because expertise in map projections and attentive observation is needed to reach a convergent result on the final projection (Balletti and Boutoura 2001, Boutoura 2006).

Nevertheless, examples exist that prove that "crowdsourcing" – i.e. the involvement of large numbers of volunteers without professional or scientific training – has been a successful strategy in other, equally challenging, domains: E.g. *Galaxy Zoo*, an online astronomy project where the general public was involved in the classification of over a million galaxies photographed by a robotic telescope (Lintott et al. 2008); or the *Australian Newspaper Digitisation Program* where volunteer users help correct errors in text extracted automatically from scanned historical newspapers (Holley 2009).

Synthesis

When analysing a historical map, it may be useful to compare aspects of it with material from other sources, or to correlate it with information from (multiple) other maps. In our vision *synthesis* would be the functionality that enables this sort of comparative analysis. It would do so by providing tools to discover related maps, annotations, and other spatial data; and merge them into one unified view, with the original historical map as the base layer.

With regard to the functionalities of our proposed system, *synthesis* is currently the least explored: We need to devise strategies to identify related annotations and map features (e.g. by matching annotation text, tags, or visual features of the annotated map fragment); develop techniques for integrating, importing and layering spatial data from external sources (e.g. vector datasets with historical country borders, raster elevation data or demographic data); and design appropriate user interfaces that support users in basic comparative analysis tasks.

Exploration

As discussed in the introduction, the vision of a tool that invites exploration of not only the map, but also of its larger historical context, was a guiding principle of our work. Consequently, our envisioned system includes a number of features and interaction metaphors to support this goal.

One, quite conventional, exploration feature is search. For any search term entered by the user, the system will perform two operations: First, it will attempt to geo-code the term (i.e. resolve the term to a geographical coordinate) using a *Gazetteer* service – a digital directory that lists place names along with their geographical location. Second, it will search for occurrences of the term in the text of earlier annotations on that map. This way, the full range of information that has been contributed to the system so far is directly available to anyone using the system.

The primary means of exploration in the system, however, is the *context tag cloud* (Figure 1): On the Web, tag clouds have been gaining popularity as a concept to compactly and efficiently visualize dominant topics and emerging key themes and trends in annotated datasets. Built from freely chosen descriptive keywords, which users have assigned to digital resources, tag clouds are especially prevalent on user-driven Web sites. Typical examples are photo sharing or social bookmarking sites, where hundreds of thousands of tagged items may be hosted and organized. Tag clouds are weighted lists that show the most often assigned tags in different font sizes, according to popularity. The resulting representation not only eases the browsing and searching of the unstructured dataset; it also helps get the "gist" of the (maybe countless) underlying items by providing a content-centric compact overview (Rivadeneira et al. 2007).

Due to these favorable properties, the use of tag clouds is increasingly investigated also in the context of geo-referenced information: One result are the so-called 'tag maps' (Jaffe et al. 2006), which combine common 2D maps with superimposed tag clouds in order to reveal important concepts of the underlying geo-tagged digital resources; or to provide an overview and filtering of unstructured spatial resources such as geo-referenced text (Slingsby et al. 2007). In a similar approach, tag clouds based on geographical resources have been proposed even for mobile usage: "Ambient Tag Clouds" (Baldauf et al. 2009) summarize digital content geo-referenced within the mobile user's proximity. Displayed on a location- (and orientation-) aware mobile device, they act as a context-sensitive exploration tool for the user's surroundings, and provide an immediate sense of "what is around" the user in the current location (and viewing direction).

Figure 1: Context tag cloud.

In essence, the intention behind our *context tag cloud* is quite similar to that of the Ambient Tag Cloud: In both cases, the goal is to provide a highly focused and condensed snapshot of only the very small fraction of the – potentially massive – geo-referenced information space that is relevant to the user. The *context tag cloud* acts as a spyglass that the user can hover over the historical map to reveal bits and pieces of knowledge collected about the map fragment underneath.

To realize the *context tag cloud*, the system needs to include functionality to accumulate, summarize and rank geo-referenced information for the map fragment the user is focusing on. Obviously, primary sources of data in this regard are the annotations collected in the system. To construct the tag cloud, the system can include not only the annotations made on the currently viewed map; it can also include annotations that users may have left on other maps (e.g. from the same epoch), at the same geographical location. Additional relevant geo-referenced content may also come from external sources: e.g. from online encyclopedias like Wikipedia[4], or community sites for sharing geo-tagged media such as Flickr[5] or Panoramio[6].

System Implementation

Although our work is still at an early stage, we have already implemented prototypes for some of the central components. The prototypes are browser-based, rich Web applications, realized with

[4] http://www.wikipedia.org/
[5] http://www.flickr.com/
[6] http://www.panoramio.com/

the *Google Web Toolkit*[7], a mature development toolkit that unifies server- and client-side Web application development in a single environment based on the Java programming language.

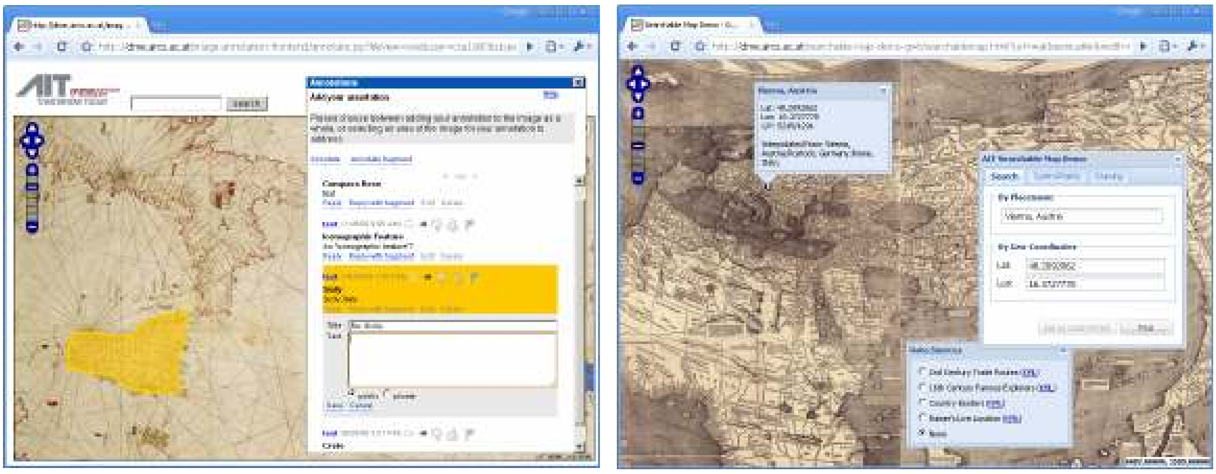

Figure 2: Map viewer prototype screenshots – phase I, annotations (left); phase II, geo-referencing and search (right).

The prototypes were implemented in three phases: In the first phase, we set up the main map browsing interface, and implemented annotation functionality. In the second phase, the focus was on a demonstrator that supports geo-referencing, based on control points and local affine transformations. Additionally, we integrated place search and basic vector data overlay (both of which require geo-referencing functionality). In the third phase, we created a proof-of-concept for the context tag cloud. The following sub-sections discuss each phase in more detail.

Implementation Phase I: Interoperable Annotations on Historical Maps

The map viewing interface we set up in the first implementation phase is based on the open source *OpenLayers*[8] Web mapping toolkit. In addition, we employed an extension produced as part of the OldMapsOnline project (Pridal and Zabicka 2008) which adds support for the *Zoomify*[9] image tile protocol.

For the implementation of the annotation functionality, we partly relied on existing work we produced for *EuropeanaConnect* project and its predecessor project *TELplus*. Within the scope of these projects, we contributed components for a comprehensive media annotation framework: in particular a browser-based annotation management interface which exchanges annotations according to the W3C *Annotea* RDF/XML protocol[10]; and a server-side 'middleware' component, which exposes a RESTful (Richardson and Ruby 2007) Web service for storing and retrieving annotations to the outside world. The middleware can be configured to work with different storage back ends such as a relational database and an RDF triple store. Additionally, it offers support to publish annotations as Linked Open Data (LOD)[11] in both human- and machine-readable form (the latter being based on RDF) to make them interoperable beyond the confines of our system (Haslhofer et al. 2009).

[7] http://code.google.com/webtoolkit/
[8] http://www.openlayers.org
[9] http://www.zoomify.com/
[10] http://www.w3.org/2001/Annotea/
[11] http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

The annotation interface already supported many of the features we needed, such as the creation of free-text annotations pertaining to the map or to map fragments defined by points, lines or polygons; editing existing annotation text or fragments; creating replies (and reply threads) to annotations; annotation 'scoping', i.e. setting an annotation's visibility to public (anyone can view the annotation) or private (only the creator can view the annotation); and a basic moderation feature that allows users to report inappropriate annotations to the system administrator via e-mail.
A screenshot of the phase I prototype is presented in Figure 2 (left): The screenshot shows the map viewer in the background, and the annotation management component as a floating window in the foreground. A single annotation is highlighted in the annotation management component; the corresponding polygon shape is overlaid on the map viewer.

Phase II: Geo-Referencing

The goal of the second implementation phase was to demonstrate geo-referencing, and some of the basic features it enables – such as place search, and import and export of vector shapes. As explained in the previous section, one way to establish a correspondence between map coordinates and a geographical coordinate system is with the help of *control points*.
Adding control points in the phase II prototype can be done iteratively, and collaboratively, just as we envision for the workbench: to seed the system with a set of initial control points, users can place markers on the map and specify their corresponding geo-coordinates. In the prototype, these can be provided either by entering them directly as WGS 84 latitude and longitude in a text entry field, or, alternatively, by typing a place name which is then automatically resolved to a geo-coordinate using the Google geo-coder service[12]. (In a future implementation, we plan to include a modern reference map to confirm or adjust the geo-coordinates of the control point. However, this is not implemented yet in the current prototype version.)
As the primary mode of how control points will be added to the system, however, we envision an iterative process based on collaborative refinement: when users search for places with the search tool, the result will almost certainly be inaccurate to some extent – either due to inaccuracies of the map itself, or due to insufficient density, inaccurate placement, or unfavorable constellation of control points. If (guided by the approximate search result) users are nonetheless able to identify the location visually, they can "correct" the search result by grabbing the result marker with their mouse and dragging it to what they consider to be the correct location. This way, the system can gather input for creating additional control points automatically in the future. In the current prototype, every corrected search result will simply create a new control point immediately.
To translate a map-image coordinate to a geo-coordinate or vice versa, the prototype computes an affine transformation based on its closest neighbor control points: Three control points are needed in order to define the coefficients of an affine transformation; however we found that we could greatly improve the robustness of our geo-referencing by selecting the four closest neighbor control points and computing four independent affine transformations (picking three control points out of the four in each run), which we then combine by taking an 'averaged median' (i.e. omitting smallest and largest result and averaging the remaining two results). Errors caused by unfavorable control point constellations (e.g. control points situated approximately along a line), in particular, could be greatly decreased this way.

[12] Google Maps API Services – Geocoding, http://code.google.com/apis/maps/documentation/geocoding

Figure 3: Geo-referencing by control points and affine transformation (world country borders overlaid for verification): mid-16th century portolan map, 4 control points (left); 1562 map of the Americas, 28 control points (right).

As direct applications of geo-referencing functionality, we added place search: Users can enter a place name in a search field, which is then resolved to a geo-coordinate using the Google geo-coder service and indicated on the map with a marker (shown in Figure 2, right image). Furthermore, we added import and overlay of vector data from KML files (e.g. see country border overlays in Figure 3).

Despite the relative simplicity of this approach, early trials have shown some promising results on various sample maps, e.g. a nautical map of the Mediterranean from the second half of the 16th century by Antonio Millo (high-resolution scan kindly provided by the Austrian National Library), or a map of the Americas from 1562 by Diego Gutiérrez (high-resolution scan available from the Library of Congress[13]), both shown in Figure 3. It is worth pointing out that, so far, our tests have been carried out with relatively few control points (4 and 28 for the examples in Figure 3, respectively); and it can be expected that results obtained in a running system, where users collaboratively add control points, will improve greatly as their density on the map grows.

Phase III: The Context Tag Cloud

To give an impression of the look-and-feel, and the general interaction mode of the context tag cloud, we integrated an early demonstrator into the user interface of the phase II prototype: As soon as the user keeps the mouse pointer idle over the map for a pre-defined period of time, the system computes a geo-coordinate for the current mouse location (based on the control points). The geo-coordinate is then resolved to a place name using the reverse geo-coding functionality of the Google geo-coder service; and the place name is used as a search query in the popular online encyclopedia Wikipedia[14]. If an article is found, tags are automatically generated from the article using a two-step process:

First, meaningful plaintext is extracted from the HTML encoded Web page. This is achieved by exploiting included HTML comments (<!-- start content --> and <!-- end content -->) to restrict the fetched markup to the actual article, removing other HTML portions such as navigation menus. In the next step, all HTML markup tags are stripped off to produce plain text. Almost

[13] http://lcweb2.loc.gov/ammem/gmdhtml/gutierrz.html
[14] http://www.wikipedia.org/

every Wikipedia article contains concluding sections for references or Web links that are irrelevant for the later text analysis and which are also removed in a last step.
Second, representative tags are generated for the resulting article plaintext by applying a simple information extraction algorithm – the *Term Frequency-Inverse Document Frequency* (TFIDF) algorithm (Salton 1989). TFIDF represents a simple method for identifying salient terms in one document out of a collection of documents (Viégas et al. 2006). The key assumption is that if a term appears frequently in one document, but not in the overall collection, it is reasonable to consider it particularly descriptive for this document.
Our implementation first splits the articles into individual terms of one, two and three words length, removing commonly used words during the process. For each term, the number of appearances is recorded (term frequency TF). For the second computation step, an index of reference documents is required. Because it is impractical to compute TFIDF against the entire collection of existing Wikipedia articles, the reference index was instead built from a general-purpose text corpus compiled from newspaper texts. (This measure also ensures that the implementation remains flexible and applicable across different topical domains.) The reference index is queried with each term; and the number of documents in which the term occurs is recorded as the document frequency (DF). The TF-IDF score is finally computed according to the following formula, where N denotes the size of the set of terms (i.e. total number of terms extracted from Wikipedia articles), and D denotes the size of the index (i.e. total number of documents indexed):

$$TFIDF = \frac{TF}{N} \cdot \log\left(\frac{D}{1+DF}\right)$$

The n highest scoring terms are then used as tags in the context tag cloud. On the client (i.e. in the user's browser), the tags are then overlaid on top of the map viewer with the help of a cross-browser open source vector drawing library[15]. A sample of the resulting output in the phase II map viewer was already shown in the previous section in Figure 1, which depicts the 16 highest scoring tags for the region name "Sardegna" (Sardinia).

Conclusion and Future Work

In this paper we presented our vision of a Web-based collaborative environment for analysing, interpreting and exploring digitised historical maps. The usage of the system we propose is governed by four basic functionalities: *annotation*, *geo-referencing*, *synthesis* and *exploration*. We proposed possible features to support these functionalities; and presented ongoing implementation work to demonstrate some of them.
In addition to completing the annotation feature set of our prototype and integrating it into the overall *EuropeanaConnect* media annotation infrastructure, we plan to direct our future effort towards deeper investigation of the context tag cloud: First and foremost, the information aggregation mechanisms, which we have described in the second section of this paper and which have not yet been demonstrated, need to be included in the prototype. User-submitted tags and free-form text, attached to map fragments in the annotation system, will be exploited as data source for the tag cloud. This can be done either by exploiting the user tags directly, or by computing a "gist" of the annotations for a particular location. The latter can be done by using a spatial variation of TFIDF that identifies terms that are salient in a certain region of the map (versus the entire map)

[15] http://raphaeljs.com/

as opposed to usual TFIDF, which identifies terms that are salient in a certain document (out of a collection of documents) (Ahern et al. 2007).

Further goals related to the context tag cloud include: the investigation of additional external sources of data that may be suitable for inclusion into the context tag cloud information base, e.g. community mapping projects such as Wikimapia[16] or Weaving History[17]; exploring how time (or historical epoch) can be used as an additional filtering criterion; and research into how the process of tag cloud generation can be made more scalable (since e.g. the application of TFIDF to a single Wikipedia article is relatively time consuming and not suitable for real-time interaction).

With regard to the geo-registration capabilities of our system, we aim: (a) to include a basic "trial-and-error" toolset that allows users to interactively test historical maps against different well-known map projections and (b) to gain a better understanding of the potential, challenges and achievable accuracy of the iterative, collaborative control point generation approach by collecting more sample data, and experimenting with additional transformation algorithms.

Acknowledgements

This paper presents work done for the best practice network *EuropeanaConnect*. *EuropeanaConnect* is a project funded by the European Commission within the area of Digital Libraries of the eContentplus Programme and lead by the Austrian National Library

References

Ahern, S., Naaman, M., Nair, R., Yang, J. H. (2007). World Explorer: Visualizing Aggregate Data From Unstructured Text in Geo-Referenced Collections. Proceedings of the 7th ACM/IEEE-CS *Joint Conference on Digital Libraries*. Vancouver, Canada: ACM.

Baldauf, M., Fröhlich, P., Reichl, P. (2009). The Ambient Tag Cloud: A New Concept for Topic-driven Mobile Urban Exploration. Proceedings of the *3rd European Conference on Ambient Intelligence*, Salzburg, Austria: Springer LNCS.

Balletti, C., Boutoura, C. (2001). Revisiting the Projective Properties of Historic Nautical Maps of the Mediterranean and the Aegean. Proceedings of the *20th International Cartographic Conference* (ICC 2001). Bejing, China.

Boutoura, C. (2006). Assigning Map Projections to Portolan Maps. *e-Perimetron, vol.1, no. 1*, 40-50.

Boutoura, C., Livieratos, E. (2006). Some Fundamentals for the Study of the Geometry of Early Maps by Comparative Methods. *e-Perimetron, vol. 1, no. 1,* 60-70.

Frisse, M. E. (1987). Searching for Information in a Hypertext Medical Handbook. *Proceedings of the ACM Conference on Hypertext* (HYPERTEXT '87), 57-66. Chapel Hill, North Carolina, United States: ACM.

Haslhofer, B., Jochum, W., King, R., Sadilek, C., Schellner, K. (2009). The Lemo Annotation Framework:Weaving Multimedia Annotations With the Web, *International Journal on Digital Libraries, vol. 10, no. 1*, 15-32.

[16] http://www.wikimapia.org/
[17] http://www.weavinghistory.org/

Holley, R. (2009). Many Hands Make Light Work: Public Collaborative OCR Text Correction in Australian Historic Newspapers. *National Library of Australia Staff Papers, ISBN 978-0-642-27694-0*.

Jaffe, A., Naaman, M., Tassa, T., Davis, M. (2006). Generating Summaries and Visualization for Large Collections of Geo-Referenced Photographs. Proceedings of the *8th ACM International Workshop on Multimedia Information Retrieval*. Santa Barbara, California, USA: ACM.

Jenny, B., Weber, A., Hurni, L. (2007). Visualizing the Planimetric Accuracy of Historical Maps with MapAnalyst. *Cartographica, vol. 42, issue 1*, 89-94.

Lintott, C. J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., Raddick, M. J., Nichol, R. C., Szalay, A., Andreescu, D., Murray, P., van den Berg, J. (2008). Galaxy Zoo: Morphologies Derived from Visual Inspection of Galaxies from the Sloan Digital Sky Survey. *MNRAS, 389, 1179*.

Pridal, P., Zabicka, P. (2008). Tiles as an approach to On-Line Publishing of Scanned Old Maps, Vedute and Other Historical Documents. *e-Perimetron, vol. 3, no. 1*.

Richardson, L., Ruby, S. (2007). *RESTful Web Services*. O'Reilly.

Rivadeneira, A. W., Gruen, D. M., Muller, M. J., Millen, D. R. (2007). Getting Our Head in the Clouds: Toward Evaluation Studies of Tagclouds. Proceedings of the *SIGCHI Conference on Human Factors in Computing Systems*. San Jose, California, USA: ACM.

Salton, G. (1989). *Automatic Text Processing – The Transformation, Analysis, and Retrieval of Information by Computer.* Addison-Wesley

Slingsby, A., Dykes, J., Wood, J., Clarke, K. (2007). Interactive Tag Maps and Tag Clouds for the Multiscale Exploration of Large Spatio-temporal Datasets. Proceedings of the *11th International Conference Information Visualization*. Zurich, Switzerland: IEEE Computer Society.

Tobler, W. R. (1966). Medieval Distortions: The Projections of Ancient Maps. *Annals of the Association of American Geographers*, Vol. 56, No. 2, 351-360.

Viégas, F. B., Golder, S., Donath, J. (2006). Visualizing Email Content: Portraying Relationships from Conversational Histories. Proceedings of the SIGCHI conference on Human Factors in Computing Systems. Montréal, Canada: ACM.